Change-point tests under local alternatives for long-range dependent processes

Johannes Tewes[*]

March 17, 2017

Abstract

We consider the change-point problem for the marginal distribution of subordinated Gaussian processes that exhibit long-range dependence. The asymptotic distributions of Kolmogorov-Smirnov- and Cramér-von Mises type statistics are investigated under local alternatives. By doing so we are able to compute the asymptotic relative efficiency of the mentioned tests and the CUSUM test. In the special case of a mean-shift in Gaussian data it is always 1. Moreover our theory covers the scenario where the Hermite rank of the underlying process changes.

In a small simulation study we show that the theoretical findings carry over to the finite sample performance of the tests..

Keywords: asymptotic relative efficiency, change-point test, empirical process, local alternatives, long-range dependence.

1 Introduction

Over the last two decades various authors have studied the change-point problem under long-range dependence and classical methods are often found to yield different results than under short-range dependence. The CUSUM test is studied in Csörgő and Horvath (1997) and compared to the Wilcoxon change-point test in Dehling et al. (2012). Ling (2007) investigates a Darling-Erdős-type result for a parametric change-point test, and estimators for the time of change are considered in Horvath and Kokoszka (1997) and Hariz et al. (2009). Moreover, the special features of long memory motivated new procedures. Beran and Terrin (1996) and Horvath and Shao (1999) are testing for a change in the linear dependence structure of the time series and Berkes et al. (2006) and Baek and Pipiras (2011) construct tests in order to discriminate between stationary long memory observations and short memory sequences with

---

[*]Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany, Email address: Johannes.Tewes@rub.de

a structural change. For a general overview of the change-point problem under long-range dependence see Kokoszka and Leipus (2001) and the associated chapter in Beran et al. (2013). One of the classical change-point problems is the change of the marginal distributions of a time series $\{Y_i\}_{i \geq 1}$. When testing for at most one change-point (AMOC) in the marginal distribution one often considers the empirical distribution function of the first k observations and that of the remaining observations. Taking a distance between the empirical distributions and the maximum over all $k < n$ yields a natural statistic. Common distances are the supremum norm, which gives the Kolmogorov-Smirnov statistic

$$T_n = \max_{1 \leq k < n} \sup_{x \in \mathbb{R}} \left| \sum_{i=1}^{k} 1_{\{Y_i \leq x\}} - \frac{k}{n} \sum_{i=1}^{n} 1_{\{Y_i \leq x\}} \right|, \tag{1.1}$$

or an L^2-distance, which gives the Cramér-von Mises statistic

$$S_n = \max_{1 \leq k < n} \int_{x \in \mathbb{R}} \left(\sum_{i=1}^{k} 1_{\{Y_i \leq x\}} - \frac{k}{n} \sum_{i=1}^{n} 1_{\{Y_i \leq x\}} \right)^2 d\hat{F}_n(x). \tag{1.2}$$

Both are widely used for goodness-of-fit tests and two-sample problems. In the change-point literature they are considered by Szyszkowicz (1994) for independent data, by Inoue (2001) for strongly mixing sequences and by Giraitis et al. (1996b) for linear long-memory processes. However, note that in the LRD setting only the Kolmogorov-Smirnov test has been investigated.

(1.1) and (1.2) are functionals of the sequential empirical process, that is $\sum_{i=1}^{\lfloor nt \rfloor} (1_{\{Y_i \leq x\}} - F(x))$ for $t \in [0, 1]$ and $x \in \mathbb{R}$. Thus the asymptotic distributions of T_n and S_n rely on that of the sequential empirical process. For weakly dependent sequences this would be a Gaussian process, in the special case of independent random variables it is called Kiefer-Müller process. For stationary sequences that exhibit long-range dependence, Dehling and Taqqu (1989a) proved that the limit process is of the form $\{J(x)Z(t)\}_{t,x}$, where $J(x)$ is a deterministic function and the process is therefore called semi-degenerate. They considered subordinated Gaussian processes, in detail $Y_i = G(X_i)$ for any measurable function G and a Gaussian sequence X_i with non-summable autocovariance function. A similar limit structure was later obtained independently by Ho and Hsing (1996) and Giraitis et al. (1996a) for long-range dependent moving-average sequences.

It is the main goal of this paper to derive the limit distribution of change-point statistics of the type (1.1) and (1.2) under local alternatives. We then apply these results to derive the asymptotic relative efficiency (ARE) of several change-point tests. To this end we investigate the sequence

$$G_1(X_1), \ldots, G_1(X_{k^*}), G_n(X_{k^*+1},) \ldots G_n(X_n), \tag{1.3}$$

Here G_n is a sequence of functions such that the distribution of $G_n(X_1)$ converges to the distribution of $G(X_1)$ in some suitable way.

Therefore, we are able to analyze various types of change-points, among them a mean-shift. Thus we may compute the ARE of Kolmogorov Smirnov, Cramér-von Mises, CUSUM and Wilcoxon test and get the surprising result that in case of Gaussian data it is always 1.

The mathematically most challenging case is the situation when the Hermite rank changes. The Hermite rank of the class $\{1_{\{G(\cdot)\leq x\}} - F(x)\}_{x\in\mathbb{R}}$ is defined as the smallest positive integer, such that $E[1_{\{G(X_1)\leq x\}}H_q(X_1)] \neq 0$ for some $x \in \mathbb{R}$, with H_q being the q-th Hermite polynomial. The structure of the limiting process $Z(t)$, e.g. the marginal distribution and the covariance structure, mainly depends on m. However, a special feature of distributional changes in subordinated Gaussian processes is the fact that the Hermite rank may change, too. Hence the question arises which Hermite process will determine the limit distribution. Under a mean-shift the Hermite rank remains unchanged, which can be seen easily by its definition. Our results differ in various ways from those obtained in Giraitis et al. (1996b), where changes in the coefficients of an LRD linear process were investigated. While the empirical process of LDR moving average sequences converges to fractional Brownian motion, we may encounter higher order Hermite processes. The possible change in the Hermite rank is therefore a novel feature in our investigation.

The rest of the paper is organized as follows. In section 2 we will state a limit theorem for the sequential empirical process under change-point alternatives. Moreover we will give the asymptotic distribution of the test statistics under the hypothesis of no change as well as under local alternatives. Thus we are able to derive the asymptotic relative efficiency of several change-point tests. In section 2.5 we consider the empirical process for long-range dependent arrays that are stationary within rows. The outcome mainly serves as a device for proving the main results, but is also of interest on its own. Section 3 contains the simulation study. To the best of our knowledge there are no results on the finite sample performance of the Cramér-von Mises change-point test under long memory. It is compared to other change-point tests and the effect of an estimated Hurst-coefficient is discussed. We obtain that the theoretical results (e.g. asymptotic relativ efficiency between Cramér-von Mises and CUSUM test) carry over to the finite sample performance of the tests. Finally proofs are provided in section 4.

2 Main results

Let $\{X_i\}_{i\geq 1}$ be a stationary Gaussian process, with

$$EX_i = 0, \quad EX_i^2 = 1 \text{ and } \rho(k) = EX_0X_k = k^{-D}L(k),$$

for $0 < D < 1$ and a slowly varying function L. The non-summability of the covariance function is one possibility to define long-range dependence. We investigate our results for so called

subordinated Gaussian processes $\{Y_i\}_{i \geq 1}$, where $Y_i = G(X_i)$ and $G \colon \mathbb{R} \to \mathbb{R}$ is a measurable function. The key tool in our analysis of possible changes in the marginal distribution of such a process is the sequential empirical process. To obtain weak convergence of this process the right normalization is given by $d_{n,m}$, defined by

$$d_{n,m}^2 = Var\left(\sum_{i=1}^n H_m(X_i)\right) \sim n^{2H} L^m(n),\tag{2.1}$$

where the constant of proportionality is $2m!(1-mD)^{-1}(2-mD)^{-1}$, see Theorem 3.1 in Taqqu (1975). $H = 1 - mD/2$ is called Hurst coefficient and

$$m = \min\left\{q > 0 \mid E[1_{\{G(X_1) \leq x\}} H_q(X_1)] \neq 0 \text{ for some } x\right\},$$

is the Hermite rank of $\{1_{\{G(\cdot) \leq x\}} - F(x)\}_{x \in \mathbb{R}}$. The mentioned result of Dehling and Taqqu (1989a) then reads as follows.

Theorem A (Dehling, Taqqu). *Let the class of functions $\{1_{\{G(\cdot) \leq x\}} - F(x)\}_{x \in \mathbb{R}}$ have Hermite rank m and let $0 < D < 1/m$. Then*

$$\frac{1}{d_{n,m}} \sum_{i=1}^{\lfloor nt \rfloor} (1_{\{G(X_i) \leq x\}} - F(x)) \xrightarrow{\mathcal{D}} \frac{J_m(x)}{m!} Z_{m,H}(t)\tag{2.2}$$

where the convergence takes place in $D([0,1] \times [-\infty, \infty])$, equipped with the uniform topology. $J_m(x)$ is defined by

$$J_m(x) = E[1_{\{G(X_1) \leq x\}} H_m(X_1)]$$

and $(Z_{m,H}(t))_{t \in [0,1]}$ is an m-th order Hermite process, see Taqqu (1979) for a definition.

Remark 2.1. In the case $m = 1$, the Hermite process becomes the well known fractional Brownian Motion, which we denote by $B_H(t)$.

2.1 The empirical process under change-point alternatives

Let us consider the following change-point model. Define the triangular array

$$Y_{n,i} = \begin{cases} G(X_i), & \text{if } i \leq \lfloor n\tau \rfloor, \\ G_n(X_i), & \text{if } i \geq \lfloor n\tau \rfloor + 1, \end{cases}\tag{2.3}$$

for measurable functions G and $(G_n)_n$ and unknown $\tau \in (0,1)$. For $\tau = 0$ one gets a row-wise stationary triangular array, as considered in section 2.5, and for $\tau = 1$ a stationary sequence, as in Dehling and Taqqu (1989a). In what follows we will denote the distribution functions of

4

$G(X_i)$ and $G_n(X_i)$ by F and $F_{(n)}$, respectively.

To obtain weak convergence of the empirical process of (2.3) we have to make some assumptions on the structure of the change and the Hermite rank.

Assumption A:

A1. The class of functions $\{1_{\{G(\cdot)\leq x\}}\}_{x\in\mathbb{R}}$ has Hermite rank m with $0 < D < 1/m$.

A2. Let $m(n)$ be the Hermite rank of $\{1_{\{G_n(\cdot)\leq x\}}\}_{x\in\mathbb{R}}$ and $m^* = \liminf_{n\to\infty} m(n)$. Then we assume

$$n^{(m-m^*)D(1+\delta)/2} \sup_{x\in\mathbb{R}}(P(\min\{G(X_1),G_n(X_1)\}\leq x) - P(\max\{G(X_1),G_n(X_1)\}\leq x)) \to 0,$$

for some $\delta > 0$.

Theorem 1. *If Assumption A holds, then*

$$\frac{1}{d_{n,m}}\sum_{i=1}^{\lfloor nt\rfloor}(1_{\{Y_{n,i}\leq x\}} - P(Y_{n,i}\leq x)) \xrightarrow{\mathcal{D}} \frac{J_m(x)}{m!}Z_{m,H}(t),$$

where $J_m(x)$ is the Hermite coefficient of $1_{\{G()\leq x\}}$. The convergence takes place in $D([0,1]\times[-\infty,\infty])$, equipped with the uniform topology.

Remark 2.2. (i) For given functions $G(x)$ and $G_n(x)$, Assumption A2 might easily being checked, see the examples below. It serves to ensure convergence of the Hermite coefficients $J_{q,n}(x) = E[1_{\{G_n(X_i)\leq x\}}H_q(X_i)]$. In detail,

$$\sup_{x\in\mathbb{R}}(P(\min\{G(X_1),G_n(X_1)\}\leq x) - P(\max\{G(X_1),G_n(X_1)\}\leq x)) \to 0$$

implies, see the proof of Lemma 4.5,

$$\sup_{x\in\mathbb{R}}|J_{q,n}(x) - J_q(x)| \to 0 \quad \forall q\in\mathbb{N}. \tag{2.4}$$

By Assumption A1, $J_1(x) = \ldots J_{m-1}(x) = 0$ for all $x\in\mathbb{R}$, yet $J_m(x)\neq 0$ for some x. Together with (2.4) this implies $m^* = \liminf_{n\to\infty} m(n) \leq m$.

(ii) Moreover, A2 implies convergence of the marginal distribution function. To see this, note

$$|F_{(n)}(x) - F(x)| = \max\{F_{(n)}(x), F(x)\} - \min\{F_{(n)}(x), F(x)\}$$
$$\leq P(\min\{G(X),G_n(X)\}\leq x) - P(\max\{G(X),G_n(X)\}\leq x)$$

and $n^{(m-m^*)D(1+\delta)/2} = O(1)$. However, the converse is not always true. Consider for instance the functions $G(x) = x$ and $G_n(x) = G_1(x) = -x$ or the situation in Example 2.8. Then again,

there are lots of natural choices of G and G_n for whom convergence of the marginal distribution functions (with a certain rate) implies Assumption A2. Among them $G_n(x) = G(x) + \mu_n$ (mean-shift), $G_n(x) = \sigma_n G(x)$ (change in variance) and

$$G_n(x) = F_{(n)}^{-1} \circ \Phi(x) \quad \text{and} \quad G(x) = F^{-1} \circ \Phi(x).$$

(iii) Our assumptions explicitly allow for the Hermite rank to change together with the marginal distribution. Then again, the limit behaivior seems to be untouched by this change. Intuitively this corresponds to the idea that the change in distribution and the change in the Hermite coefficient, both caused by the difference of G and G_n, are of the same order. For $q < m$ this enforces the function $J_{q,n}(x)$ to converge rather fast to 0. Technically this can be explained through A2. If this assumption is dropped, we might actually encounter limits with multiple Hermite processes. Such cases will be considered in Example 2.8 and Corollary 2.13.

(iv) If A1 is violated, the sequence $\{G(X_i)\}_{i \geq 1}$ is actually short-range dependent. For stationary observations Csörgő and Mielniczuk (1996) showed convergence of the sequential empirical process to a two-parameter Gaussian process. Change-point alternatives have not been considered for such random variables, yet, but would require fundamentally different proofs compared to our results.

2.2 Asymptotic behavior of the change-point statistics

We now apply the results concerning empirical processes to determine the asymptotic distribution of the Kolmogorov-Smirnov statistics

$$T_n = \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} d_{n,m}^{-1} \left| \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right| \tag{2.5}$$

and that of the Cramér-von Mises change-point statistic

$$S_n = d_{n,m}^{-2} \sup_{t \in [0,1]} \int_{\mathbb{R}} \left| \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right|^2 d\hat{F}_n(x). \tag{2.6}$$

To get a non degenerate limit under a sequence of local alternatives it is important to choose the right amount of change. For a mean-shift this is naturally the difference of the expectations before and after the change. For a general change we formulate the test problem as follows: We wish to test the hypothesis

$$\mathbf{H}: \quad \text{Assumption A1 holds and } G_n(x) = G(x) \text{ for all } x \in \mathbb{R} \text{ and } n \geq 1,$$

against the sequence of local alternatives

$\mathbf{A_n}$: Assumption A holds and, for $n \to \infty$,

$$\frac{n}{d_{n,m}}(F(x) - F_{(n)}(x)) \to g(x), \tag{2.7}$$

uniformly in x, where $g(x)$ is a measurable function of bounded

total variation, whose support has positive Lebesgue measure.

Remark 2.3. Note that $nd_{n,m}^{-1} \sim n^{mD/2}L^{-m/2}(n)$. Thus (2.7) implies

$$n^{(m-m^*)D(1+\delta)/2}(F(x) - F_{(n)}(x)) \to 0,$$

for $\delta < m^*/(m - m^*)$ or $m^* = m$. This again implies Assumption A2 for certain choices of functions G and G_n, see Remark 2.2 (ii).

Theorem 2. *(i) Under the hypothesis* $\mathbf{H}$ *of no change we have, as* $n \to \infty$,

$$T_n \xrightarrow{\mathcal{D}} \sup_{x \in \mathbb{R}} |J_m(x)/(m!)| \sup_{t \in [0,1]} \left| \tilde{Z}_{m,H}(t) \right|$$

$$and \quad S_n \xrightarrow{\mathcal{D}} \int_{x \in \mathbb{R}} (J_m(x)/(m!))^2 \, dF(x) \sup_{t \in [0,1]} \left| \tilde{Z}_{m,H}(t) \right|^2,$$

where $\tilde{Z}_{m,H}(t) = Z_{m,H}(t) - tZ_{m,H}(1)$.
(ii) Under the sequence of local alternatives $\mathbf{A_n}$ *we have, as* $n \to \infty$,

$$T_n \xrightarrow{\mathcal{D}} \sup_{x \in \mathbb{R}} \sup_{t \in [0,1]} \left| J_m(x)/(m!)\tilde{Z}_{m,H}(t) - g(x)\psi_\tau(t) \right|$$

$$and \quad S_n \xrightarrow{\mathcal{D}} \sup_{t \in [0,1]} \int_{x \in \mathbb{R}} \left(J_m(x)/(m!)\tilde{Z}_{m,H}(t) - g(x)\psi_\tau(t) \right)^2 \, dF(x),$$

where

$$\psi_\tau(t) = \begin{cases} t(1 - \tau), & \text{if } t \leq \tau, \\ \tau(1 - t), & \text{if } t > \tau. \end{cases}$$

Motivated by this Theorem we consider change-point tests based on the statistics T_n and S_n. Critical values might be chosen as

$$\sup_{x \in \mathbb{R}} |J_m(x)/(m!)|q_{1-\alpha,m,H} \quad \text{and} \quad \int_{x \in \mathbb{R}} (J_m(x)/(m!))^2 \, dF(x)q_{1-\alpha,m,H}^2,$$

for the Kolmogorov-Smirnov test and the Cramér-von Mises test, respectively. Here $q_{1-\alpha,m,h}$ is the $(1-\alpha)$-quantile of $\sup_{t \in [0,1]} |\tilde{Z}_{m,H}(t)|$. Thereby the tests have asymptotically level α and nontrivial power against local alternatives.

The tests can be performed, if the right normalization for the empirical process, the supremum of $J_m(x)$ and the distribution of $\sup_{t \in [0,1]} |\tilde{Z}_{m,H}(t)|$ are known. In practical applications this might be not the case. Solutions are self-normalization (Shao (2011)), estimating the the Hurst-coefficient (see for example Künsch (1987)) and bootstrap estimators for $J_m(x)$ (Tewes (2016)).

2.3 Examples

Example 2.4 (Mean-shift)**.** Let $G_n(x) = G(x) + \mu_n$ with $\mu_n \sim d_n/n$, then we get the typical change in the mean problem. In the case of long-range dependent subordinated Gaussian processes this was considered in Dehling et al. (2012, 2013), Csörgő and Horvath (1997), Shao (2011) and Betken (2016). Let f_G be the probability density of $G(X_1)$, and assume that it is continuous and of bounded variation. Then we obtain

$$\frac{n}{d_{n,m}}(F(x) - F_{(n)}(x)) = \frac{n}{d_{n,m}}(F(x) - F(x - \mu_n)) \to C f_G(x),$$

where, due to continuity of f_G, the convergence holds uniformly.

Example 2.5 (Change in the variance)**.** To describe the change-in-variance-problem define $G_n(x) = 1/(1 - \delta_n)G(x)$, with $\delta_n \sim d_n/n$. For ease of notation let $\delta_n = d_n/n$. Then we get

$$\sup_{x \in \mathbb{R}} \left| \delta_n^{-1}(F(x) - F_{(n)}(x)) - x f_G(x) \right|$$

$$= \sup_{x \in \mathbb{R}} \left| \delta_n^{-1}(F(x) - F(x - \delta_n x)) - x f_G(x) \right|$$

$$= \sup_{x \in \mathbb{R}} \left| \frac{xF(x) - (x - \delta_n x)F(x - \delta_n x)}{\delta_n x} - F(x - \delta_n x) - x f_G(x) \right|$$

$$\leq \sup_{x \in \mathbb{R}} \left| \frac{xF(x) - (x - \delta_n x)F(x - \delta_n x)}{\delta_n x} - (x f_G(x) + F(x)) \right| \tag{2.8}$$

$$+ \sup_{x \in \mathbb{R}} |F(x - \delta_n x) - F(x)|. \tag{2.9}$$

The derivative of $xF(x)$ is $x f_G(x) + F(x)$, hence (2.8) converges to 0. The convergence is uniform, if f_G and F are continuous. (2.9) converges to 0, because of continuity, monotonicity and boundedness of F. Thus (2.7) holds with function $g(x) = x f_G(x)$. Assume without loss of generality $\sigma_n = 1/(1 - \delta_n) > 1$, then

$$P \max\{G(X_1), G_n(X_1)\} \leq x)$$

$$= P(\sigma_n G(X_1) \leq, G(x) \geq 0) + P(G(X_1) \leq x, G(X_1) \leq 0)$$

$$= \begin{cases} F(x/\sigma_n), & \text{if } x \geq 0, \\ F(x), & \text{if } x < 0. \end{cases}$$

The minimum can be treated analogously, hence Assumption A2 follows from convergence of the marginals.

Additionally one might consider a combined change in mean and variance, given through $G_n(x) = \sigma_n G(x) + \mu_n$. In this case (2.7) holds with $g(x) = f_G(x)(C_1 + C_2 x)$.

Example 2.6 (Generalized inverse of a mixture distribution)**.** By using the generalized inverse of a distribution function one could generate subordinated Gaussian processes with any given marginals, see for example Dehling et al. (2013). We use this for the change-point problem by setting

$$G \equiv F^{-1} \circ \Phi \quad \text{and} \quad G_n \equiv F_{(n)}^{-1} \circ \Phi.$$

For a continuous distribution function F^* define the mixture

$$F_{(n)}(x) = (1 - \delta_n)F(x) + \delta_n F^*(x),$$

with $\delta_n \sim d_n n^{-1}$. Then (2.7) holds with $g(x) = F^*(x) - F(x)$ and moreover

$$\begin{aligned}
P(\max\{G(X_1), G_n(X_1)\} \leq x) &= P(\max\{F^{-1} \circ \Phi(X_1), F_{(n)}^{-1} \circ \Phi(X_1)\} \leq x) \\
&= P(\Phi(X_1) \leq \min\{F(x), F_{(n)}(x)\}) \\
&= \min\{F(x), F_{(n)}(x)\}.
\end{aligned}$$

Analogously one has $P(\min\{G(X_1), G_n(X_1)\} \leq x) = \max\{F(x), F_{(n)}(x)\}$. Hence

$$P(\min\{G(X_1), G_n(X_1)\} \leq x) - P(\max\{G(X_1), G_n(X_1)\} \leq x) = |F_{(n)}(x) - F(x)|,$$

thus Assumption A2 is also satisfied. For strongly mixing data similar local alternatives were considered by Inoue (2001).

Example 2.7 (χ^2-distribution)**.** Consider a χ^2-distribution given through $G(x) = x^2$ and note that the indicator functions have Hermite rank $m = 2$, see also Dehling and Taqqu (1989a). Further let

$$G_n(x) = \begin{cases} a_n x^2, & \text{if } x \geq 0, \\ x^2, & \text{if } x < 0, \end{cases}$$

with Hermite ranks $m(n) = 1$ for all $n \in \mathbb{N}$. If $(a_n - 1) \sim d_{n,2}/n$, then one can show (similar

to the case of a variance change in Example 2.5) that

$$\frac{n}{d_{n,2}} \left(P\left(G(X_1) \le x \right) - P\left(G_n(X_1) \le x \right) \right)$$

$$\to C\sqrt{x}\phi(\sqrt{x})1_{[0,\infty)}(x),$$

uniformly in x. As Assumption A2 is satisfied, too, we may apply Corollary 2 (ii) with function $g(x) = C\sqrt{x}\phi(\sqrt{x})1_{[0,\infty)}(x)$ and $m = 2$.

Example 2.8 (Multiple Hermite processes in the limit)**.** In the previous example, together with the marginal distribution, also the Hermite rank has changed. However, the limiting process seems to be untouched by this fact and one might ask whether this is intuitive or not. It is caused by the fact that the change in the distribution and the change in the Hermite coefficients, both originating in the difference of the functions $G(x)$ and $G_n(x)$, are of the same order.

To get an additional Hermite process in the limit, one would need $(a_n - 1) \sim d_{n,2}/d_{n,1}$, see Corollary 2.13 and its proof. But then

$$\frac{n}{d_{n,2}} \sup_x \left| F(x) - F_{(n)}(x) \right| = \frac{n}{d_{n,1}} \frac{d_{n,1}}{d_{n,2}} \sup_x \left| F(x) - F_{(n)}(x) \right| \to \infty,$$

and the test would have asymptotic power 1.

To achieve nontrivial asymptotic power one has to consider structural breaks that consists of two aspects and where only one is captured by the marginal distribution. To this end define the transformations

$$G(x) = \Phi^{-1}(F(|x|)) = \Phi^{-1}(2\Phi(|x|) - 1)$$

and

$$G_n(x) = \Phi^{-1}(F_{(n)}^*(G_n^*(x))) + \mu_n,$$

where $F_{(n)}^*(x) = P(G_n^*(X_i) \le x)$ and

$$G_n^*(x) = \begin{cases} a_n x^2, & \text{if } x \ge 0, \\ x^2, & \text{if } x < 0, \end{cases}$$

for some sequence $(a_n)_n$ with $a_n \ne 1$ and $a_n \to 1$. On the one hand, $\{1_{\{G() \le x\}}\}_x$ has Hermite rank $m = 2$ and $G(X_i) \sim N(0,1)$. On the other hand, $\{1_{\{G_n() \le x\}}\}_x$ has Hermite rank $m(n) = 1$ for all $n \in \mathbb{N}$ and $G_n(X_i) \sim N(\mu_n, 1)$. Now let $\mu_n \sim d_{n,2}/n$, then Example 2.4 applies and we

obtain

$$\frac{n}{d_{n,2}}(F_{(n)}(x) - F(x)) = \frac{n}{d_{n,2}}(\Phi(x - \mu_n) - \Phi(x)) \to C\phi(x),$$

for any sequence $(a_n)_n$. In contrast, the convergence of the Hermite coefficients is highly influenced by $(a_n)_n$. If the sequence is chosen such that $(a_n - 1) \sim d_{n,2}/d_{n,1}$ (therefore, it converges slower than μ_n), then the sequential empirical process will converge towards

$$K(x,t) = \begin{cases} J_2(x)/2Z_2(t), & \text{if } t \leq \tau, \\ \tilde{J}_1(x)Z_1(t) + J_2(x)/2Z_2(t)+, & \text{if } t > \tau. \end{cases}$$

Actually this can be proved similar to Corollary 2.13. Moreover, the Kolmogorov-Smirnov statistic converges weakly to

$$\sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |K(x,t) - tK(x,1) - \psi_\tau(t)C\phi(x)|.$$

We find this example rather pathological, therefore such situations are excluded from the main results via Assumption A2.

2.4 Asymptotic relative efficiency

By studying the asymptotic distributions under local alternatives one might compare different tests in terms of the asymptotic relative efficiency (ARE). Here we give a precise definition of the ARE in the very special context of our change-point setting. The general idea is due to Pitman (1948) (for a published article see for example Noether (1950)) and was formalized in Noether (1955). Of course it can be extended to all kinds of testing procedures.

Definition 2.9. Let T_1 and T_2 represent two change-point test procedures. Consider the local alternatives

$$(G, G_{n_k}, \tau) \text{ and a sample size } (n_k)_k,$$
$$(G, \tilde{G}_{m_k}, \tau) \text{ and a sample size } (m_k)_k,$$

such that $G_{n_k}(x) = \tilde{G}_{m_k}(x) = G_k(x)$ for all $k \geq 1$ and $x \in \mathbb{R}$.
Let β_1 be the asymptotic power of the test T_1 against the local alternatives given by $(G, G_{n_k}, \tau, (n_k)_k)$ and β_2 be the asymptotic power of the test T_2 against the local alternatives given by $(G, \tilde{G}_{m_k}, \tau, (m_k)_k)$. If β_1 equals β_2, then the *asymptotic relative efficiency* (ARE) of the tests T_1 and T_2 is defined as

$$ARE(T_1, T_2) = \lim_{k \to \infty} \frac{m_k}{n_k}.$$

Example 2.10 (Mean-shift in Gaussian data). Consider $G(x) = x$ and $G_n(x) = G(x) + \mu_n$, in other words a mean-shift in Gaussian data. As for the Hermite coefficient function, we get $J_1(x) = -\phi(x)$, where ϕ is the standard normal probability density. Thus, according to Corollary 2, the test statistic T_n converges towards

$$\sup_{x\in\mathbb{R}}|\phi(x)| \sup_{t\in[0,1]}\left|\tilde{B}_H(t) + C\psi_\tau(t)\right| = (2\pi)^{-1/2} \sup_{t\in[0,1]}\left|\tilde{B}_H(t) + C\psi_\tau(t)\right|,$$

whereas under the Null, that is we have a stationary standard Gaussian sequence, the limit distribution would be

$$\sup_{x\in\mathbb{R}}|\phi(x)| \sup_{t\in[0,1]}\left|\tilde{B}_H(t)\right| = (2\pi)^{-1/2} \sup_{t\in[0,1]}\left|\tilde{B}_H(t)\right|.$$

For the Cramér-von Mises statistic we obtain analogously the limit distributions

$$\int \phi^3(x)dx \sup_{t\in[0,1]}\left|\tilde{B}_H(t) + C\psi_\tau(t)\right|^2 \quad \text{and} \quad \int \phi^3(x)dx, \ \sup_{t\in[0,1]}\left|\tilde{B}_H(t)\right|^2$$

under local alternative and hypothesis, respectively. Hence in this special case the CUSUM test, the Wilcoxon test (see Dehling et al. (2013) for each), the Kolmogorov-Smirnov test and the Cramér-von Mises test all have the same asymptotic power, namely

$$P\left(\sup_{t\in[0,1]}|\tilde{B}_H(t) + C\psi_\tau(t)| > q_{1-\alpha,H}\right), \tag{2.10}$$

where $q_{1-\alpha,H}$ is the $(1-\alpha)$-quantile of the maximum of a fractional Brownian bridge $\sup_{t\in[0,1]}|\tilde{B}_H(t)|$. As a direct consequence, one gets that the ARE of the four tests is 1. This result is quite surprising, keeping in mind that CUSUM and Wilcoxon tests are designed to detect level-shifts, while our tests have power against all kinds of distributional changes.

For non-Gaussian data and change-points beyond a simple mean-shift, the investigation of the ARE is not that straightforward. In fact, little is known about the distribution of

$$\sup_t|\tilde{B}_H(t) + f(t)|,$$

and even less if higher order Hermite processes are considered. This seems to prevent a precise computation of the ARE in many cases. However, one might derive lower bounds for the efficiency as we do in next example for a combined change in mean and variance. Unlike in the previous example, we will make use of the subtle definition of the ARE.

Example 2.11 (Combined change in mean and variance). Let $G(x) = x$ and $G_k(x) = \sigma_k x + \mu_k$, that is a combined change of mean and variance in Gaussian data. If further $\mu_k k/d_{k,1} \to C_1 > 0$ and $(1 - 1/\sigma_k)k/d_{k,1} \to C_2 > 0$, then by example 2.5 the empirical bridge-type process

converges to (for the sample size $k \to \infty$)

$$\phi(x)\tilde{B}_H(t) + \phi(x)(C_1 + C_2 x)\psi_\tau(t), \quad x \in \mathbb{R}, \ t \in [0,1].$$

We now consider slightly modified Cramér-von Mises and CUSUM tests, in detail, instead of $[0,1]$ the supremum is taken over $[\kappa_1, \kappa_2]$ for some $\kappa_1 \in (0, 1/2)$ and $\kappa_2 \in (1/2, 1)$.

The asymptotic distribution of the CUSUM test has been derived in Dehling et al. (2013), but only in the case of a mean-shift with constant variance. However, for $EG^2(X_i) < \infty$ the CUSUM statistic is a continuous functional of the sequential empirical process. Thus, we might apply our Theorem 1 and conclude that the CUSUM statistic converges under this type of local alternatives to

$$\sup_{t \in [\kappa_1, \kappa_2]} \left| \int \phi(x) \left(\tilde{B}_H(t) + (C_1 + C_2 x)\psi_\tau(t) \right) dx \right|$$
$$= \sup_{t \in [\kappa_1, \kappa_2]} \left| \tilde{B}_H(t) + C_1 \psi_\tau(t) \right|.$$

Note that this is the same limit as under a mean-shift with constant variance and thus, too, the asymptotic power is the same as in example 2.10.

The limiting distribution of the Cramér-von Mises statistic is given by

$$Z^2 = \sup_{t \in [\kappa_1, \kappa_2]} \int \phi^3(x) \left(\tilde{B}_H(t) + (C_1 + C_2 x)\psi_\tau(t) \right)^2 dx$$
$$= \sup_{t \in [\kappa_1, \kappa_2]} \left\{ \int \phi^3(x) dx \left(\tilde{B}_H(t) + C_1 \psi_\tau(t) \right)^2 + C_2^2 \psi_\tau^2(t) \int \phi^3(x) x^2 dx \right\},$$

and for its asymptotic power we obtain

$$P\left(Z^2 > q_{1-\alpha,H}^2 \int \phi^3(x) \, dx \right)$$
$$= P\left(Z^2 > q_{1-\alpha,H}^2 \int \phi^3(x) \, dx \, , \ \sup_{t \in [\kappa_1, \kappa_2]} \{\tilde{B}_H(t)\} > q_{1-\alpha,H} \right) \quad (2.11)$$
$$+ P\left(Z^2 > q_{1-\alpha,H}^2 \int \phi^3(x) \, dx \, , \ \sup_{t \in [\kappa_1, \kappa_2]} \{\tilde{B}_H(t)\} \leq q_{1-\alpha,H} \right).$$

First assume $\sup_t\{\tilde{B}_H(t)\} \leq q = q_{1-\alpha,H}$ and consider C_1^*, given by

$$C_1^* = f^*(C_1, C_2, q, \tau, \kappa_1, \kappa_2)$$
$$= \min_{t \in [\kappa_1, \kappa_2]} \left\{ \frac{\sqrt{q^2 + 2q C_1 \psi_\tau(t) + \left(C_1^2 + C_2^2(\int \psi^3(x) x^2 dx / \int \phi^3(x) dx) \right) \psi_\tau^2(t)} - q}{\psi_\tau(t)} \right\}.$$

13

Now C_1^* is constructed in a way, such that [1]

$$C_1^* > C_1$$

and for all $\omega \in \Omega$ with $\sup_t \tilde{B}_H(t;\omega) \leq q$

$$
\begin{aligned}
Z^2 &= \sup_{t \in [\kappa_1, \kappa_2]} \left\{ \int \phi^3(x) dx \left(\tilde{B}_H(t) + C_1 \psi_\tau(t) \right)^2 + C_2 \psi_\tau^2(t) \int \phi^3(x) x^2 dx \right\} \\
&> \sup_{t \in [\kappa_1, \kappa_2]} \left\{ \int \phi^3(x) dx \left(\tilde{B}_H(t) + C_1^* \psi_\tau(t) \right)^2 \right\}.
\end{aligned}
$$

If, on the other hand, $\sup_t \{ \tilde{B}_H(t) \} > q_{1-\alpha,H}$, then (because $C_1 > 0$) automatically $Z^2 > q_{1-\alpha,H}^2 \int \phi^3(x) dx$. Combining these two findings with (2.11) we can bound the asymptotic power from below by

$$
\begin{aligned}
&P \left(Z^2 > q_{1-\alpha,H}^2 \int \phi^3(x)\, dx \right) \\
&= P \left(\sup_{t \in [\kappa_1, \kappa_2]} \int \phi^3(x) \left(\tilde{B}_H(t) + C_1^* \psi_\tau(t) \right)^2\, dx > q_{1-\alpha,H}^2 \int \phi^3(x) dx \right) \\
&\geq P \left(\sup_{t \in [\kappa_1, \kappa_2]} \left| \tilde{B}_H(t) + C_1^* \psi_\tau(t) \right| > q_{1-\alpha,H} \right),
\end{aligned}
\tag{2.12}
$$

for $C_1^* > C_1$.

Now we are ready to compute the ARE. To this end we chose different sample sizes for both test. In detail, $(n_k)_k$ for the Cramér-von Mises test and $(m_k)_k$ for the CUSUM test. Moreover, the local alternatives are such that $G_{n_k}^{CvM}(x) = G_{m_k}^{CUSUM}(x) = G_k(x)$ for all $x \in \mathbb{R}$, consequently

$$\mu_{n_k}^{(1)} = \mu_{m_k}^{(2)} = \mu_k \quad \text{and} \quad \sigma_{n_k}^{(1)} = \sigma_{m_k}^{(2)} = \sigma_k.$$

For the CUSUM test, in order to achieve at least asymptotic power β, its limit distribution has to satisfy

$$P \left(\sup_{t \in [0,1]} |\tilde{B}_H(t) + C_1^* \psi_\tau(t)| > q_{\alpha,H} \right) \geq \beta.$$

In other words, $C_1^* = \pi^{-1}(\beta)$, where

$$\pi(C^*) = P \left(\sup_{t \in [0,1]} |\tilde{B}_H(t) + C_1^* \psi_\tau(t)| > q_{\alpha,H} \right)$$

---

[1] Here weed the restriction to $[\kappa_1, \kappa_2]$.

and π^{-1} is the generalized inverse. Therefore, the sample size of the CUSUM test has to be chosen such that

$$C_1^* = \lim_{k \to \infty} \frac{m_k}{d_{m_k}} \mu_{m_k}^{(2)} = \lim_{k \to \infty} \frac{m_k}{d_{m_k}} \mu_k. \tag{2.13}$$

We also obtain (as μ_k and $(1 - 1/\sigma_k)$ are of the same order)

$$\lim_{k \to \infty} \frac{m_k}{d_{m_k}} \left(1 - \frac{1}{\sigma_{m_k}^{(2)}} \right) = \lim_{k \to \infty} \frac{m_k}{d_{m_k}} \left(1 - \frac{1}{\sigma_k} \right) = C_2^*,$$

for some $C_2^* > 0$.

Next we will select the sample size for the Cramér-von Mises test such that its asymptotic power might be bounded from below as in (2.12) (and therefore by β). To this end choose $C_1 > 0$, such that

$$\tilde{f}(C_1) = f\left(C_1, C_1 \frac{C_2^*}{C_1^*}, q, \tau, \kappa_1, \kappa_2 \right) = C_1^*.$$

The function $\tilde{f} \colon [0, \infty) \to [0, \infty)$ is monotone increasing, surjective and continuous (as the minimum is attained either in κ_1 or κ_2), therefore such an C_1 always can be found. By construction of the function f it follows that $C_1 < C_1^*$. Now let the sample size of the Cramér-von Mises test satisfy

$$C_1 = \lim_{k \to \infty} \frac{n_k}{d_{n_k}} \mu_{n_k}^{(1)} = \lim_{k \to \infty} \frac{n_k}{d_{n_k}} \mu_k. \tag{2.14}$$

Moreover, we observe

$$
\begin{aligned}
C_2 &= \lim_{k \to \infty} n_k / d_{n_k} (1 - 1/\sigma_k) \\
&= \lim_{k \to \infty} (n_k / d_{n_k} \mu_k) \left(m_k / d_{m_k} \mu_k \right)^{-1} \left(m_k / d_{m_k} (1 - 1/\sigma_k) \right) \\
&= C_1 C_2^* / C_1^*.
\end{aligned}
$$

Thus,

$$f(C_1, C_2, q, \tau, \kappa_1, \kappa_2) = C_1^*,$$

15

and we observe for the asymptotic power of the Cramér-von Mises test

$$P \left(\sup_{t \in [\kappa_1, \kappa_2]} \left\{ \int \phi^3(x) \, dx \left(\tilde{B}_H(t) + C_1 \psi_\tau(t) \right)^2 + C_2 \psi_\tau^2(t) \int \phi^3(x) x^2 \, dx \right\} > q_{1-\alpha,H} \int \phi^3(x) dx \right)$$

$$\geq P \left(\sup_{t \in [\kappa_1, \kappa_2]} \left| \tilde{B}_H(t) + C_1^* \psi_\tau(t) \right| > q_{1-\alpha,H} \right)$$

$$\geq \beta.$$

So both test have (at least) asymptotic power β against the local alternatives (G, G_k, τ). Finally,

$$\left(\frac{m_k}{n_k} \right)^{1-H} = \frac{m_k}{d_{m_k}} \mu_k \frac{d_{n_k}}{n_k} \mu_k^{-1} \frac{L^{(1/2)}(m_k)}{L^{(1/2)}(n_k)}$$
$$\to \frac{C_1^*}{C_1} > 1,$$

by construction of the sample sizes and the definition of slowly varying functions. Consequently

$$ARE(CvM, CUSUM) = (C_1^*/C_1)^{1/(1-H)} > 1.$$

In other words, the Cramér-von Mises test is asymptotically more efficient, no matter how small the additional variance-change is.

2.5 The empirical process of triangular arrays

Since the work of Dehling and Taqqu (1989a,b), uniform reduction principles have become the main tool in the analysis of empirical processes of long-range dependent data. More precisely, the empirical process gets approximated only by the first term of its Hermite expansion (if the underlying process is not Gaussian other expansions are available). However, most results are investigated for stationary sequences. When considering $G(X_1), \ldots, G(X_{\lfloor n\tau \rfloor}), G_n(X_{\lfloor n\tau \rfloor+1}), \ldots, G_n(X_n)$, the empirical process of the first $\lfloor n\tau \rfloor$ random variables can be approximated just as in Dehling and Taqqu (1989a). In contrast the Hermite expansion of $1_{\{G_n(X_i) \leq x\}} - F_{(n)}(x)$ is

$$\sum_{q=m^*}^{\infty} \frac{J_{q,n}(x)}{q!} H_q(X_i).$$

Two difficulties arise. First, m^* might be smaller than m, the Hermite rank of $\{1_{\{G(\cdot) \leq x\}}\}_{x \in \mathbb{R}}$. Secondly, the coefficients $J_{q,n}(x)$ depend on n and might converge uniformly to 0. Thus, it is a priori not clear which term of the Hermite expansion is asymptotically dominant or if there are even more than one. The next result is a reduction principle that lays emphasis on this aspects. We will make use of it in the proof of Theorem 1, but is also of interest on its own.

Theorem 3. *Let $\{G_n\}_n$ be a sequence of measurable functions and let $m(n)$ be the sequence of Hermite ranks of $\{1_{\{G_n(\cdot)\leq x\}}\}_{x\in\mathbb{R}}$. Then, for any $m \in \mathbb{N}$ with $m(n) \leq m < 1/D$ (for $n \geq n_0$),*

$$
P\left(\sup_{t\in(0,1)}\sup_{x\in\mathbb{R}}\frac{1}{d_{n,m}}\left|\sum_{i=1}^{\lfloor nt\rfloor}(1_{\{G_n(X_i)\leq x\}}-\sum_{q=0}^{m}\frac{J_{q,n}(x)}{q!}H_q(X_i))\right|>\epsilon\right)\leq Cn^{-\kappa}(1+\epsilon^{-3}),
$$

where C and κ do not depend on n.

Remark 2.12. (i) Theorem 3 contains the reduction principle of Dehling and Taqqu (1989a) as a special case ($G_n(x) = G(x)$ and $m(n) = m$).

(ii) Note that $\{1_{\{G_n(\cdot)\leq x\}}-F_{(n)}(x)\}_{x\in\mathbb{R}}$ might has a Hermite rank smaller than m (say $m^* < m$). Thus, one might expect d_{n,m^*}^{-1} as normalization. The weaker normalization $d_{n,m}^{-1}$ is however possible since the empirical process is approximated by additional terms of the Hermite expansion, in detail those up to m.

(iii) A similar result is given by Wu (2003), who considers linear long memory processes and even shows convergences with respect to a weighted supremum metric. Then again, he considers only the normal empirical process, while we also treat the sequential version. Moreover, we consider triangular arrays, which Wu (2003) does not.

Corollary 2.13. *Let $\{G_n\}_n$ be sequence of measurable functions and let $m(n)$ be the sequence of Hermite ranks of $\{1_{\{G_n(\cdot)\leq x\}}\}_{x\in\mathbb{R}}$. If further $m^* \leq m(n) \leq m < 1/D$ for all $n \geq n_0$ and*

$$
\frac{d_{n,q}}{d_{n,m}}\frac{J_{q,n}(x)}{q!} \to h_q(x) \quad \forall q \in \{m^*,\ldots,m\},
$$

uniformly in x, then

$$
\frac{1}{d_{n,m}}\sum_{i=1}^{\lfloor nt\rfloor}(1_{\{G_n(X_i)\leq x\}}-F_{(n)}(x)) \xrightarrow{\mathcal{D}} \sum_{q=m^*}^{m}h_q(x)Z_q(t).
$$

$(Z_{q,H}(t))_{t\in[0,1]}$ are uncorrelated, qth order Hermite processes.

Remark 2.14. (i) Comparing the limit process of Corollary 2.13 to that of Theorem 1 it is apparent that multiple Hermite Processes are involved. This is not the case in Theorem 1. The reason is Assumption A2, which causes the Hermite coefficients $J_{m,n}(x)$ to converge rather fast.

(ii) The Hermite processes occurring in the limit are dependent, see Proposition 1 in Bai and Taqqu (2013).

Remark 2.15. In view of the proof of Corollary 2.13 it is important to note that the functions h_q are uniform limits of the cádlág-functions $J_{m,n}(x)$ and hence elements of $D[-\infty,\infty]$. As a consequence they are also bounded (Pollard (1984)).

Example 2.16. There are indeed sequences of functions $\{G_n\}_n$ that satisfy the conditions of Corollary 2.13. Consider again the functions from Example 2.7, namely $G_n(x) = x^2(1_{x \geq 0} + a_n 1_{x<0})$ with $a_n \to 1$ and $a_n \neq 1$. Thus, we are in the situation of Theorem 3 with $m(n) = 1$ for all $n \in \mathbb{N}$. One obtains, $a_n \to 1$s

$$\sup_{x \in \mathbb{R}} |J_{2,n}(x) - J_2(x)| \to 0,$$

with

$$J_2(x) = E[1_{\{X_1^2 \leq x\}}(X_1^2 - 1)] = -2\sqrt{x}\phi(\sqrt{x})1_{\{x \geq 0\}}.$$

If in addition $a_n \sim n^{-D/2}L^{1/2}(n) \sim d_{n,2}/d_{n,1}$, then

$$\sup_x \left| \frac{d_{n,1}}{d_{n,2}} J_{1,n}(x) - Cx\phi(\sqrt{x})1_{x \geq 0} \right| \to 0,$$

for some constant C depending on D only. Corollary 2.13 then holds with $m = 2$, $m^* = 1$, $h_1(x) = Cx\phi(\sqrt{x})1_{x \geq 0}$ and $h_2(x) = J_2(x)/2$.

3   Simulation Study

3.1   Fractional Gaussian Noise

Consider a mean-shift in Gaussian data. Then Example 2.4 states that the Cramér-von Mises test (and the Kolmogorov-Smirnov test) are asymptotically as efficient as the CUSUM test. The goal of this simulation study is to examine whether this theoretical and asymptotic result carry over to the finite sample performance of the tests. We will consider samples of size 50 to 400. For these situations the approximation of the empirical process by its semi-degenerate limit process is quite inaccurate. The empirical size of the Cramér-von Mises test will be therefore much larger than the nominal size, if critical values are deduced from the asymptotic distribution. Instead we simulate $J = 1000$ Gaussian time series

$$X_{j,1}, \ldots X_{j,n} \quad j = 1, \ldots J$$

with Hurst coefficient H. In the simulation study we will use *fractional Gaussian noise* for this sequences. Subsequently, a Cramér-von Mises statistic is calculated for each of the $J = 1000$ Gaussian series, in detail

$$S_{n,j} = \max_{1 \leq k < n} \int_{x \in \mathbb{R}} \left(\sum_{i=1}^{k} 1_{\{X_{j,i} \leq x\}} - \frac{k}{n} \sum_{i=1}^{n} 1_{\{X_{j,i} \leq x\}} \right)^2 d\hat{F}_{n,j}(x) \quad j = 1, \ldots J.$$

Table 1: Empirical power, H assumed to be known, size of level shift $\mu = 1$, relative change positions $\tau = 0.2$ and $\tau = 0.5$.

	Relative change position $\tau = 0.2$							
	$H = 0.6$				$H = 0.7$			
n	50	100	250	400	50	100	250	400
S	0.196	0.566	0.854	0.970	0.158	0.215	0.547	0.689
W	0.263	0.525	0.910	0.983	0.201	0.233	0.501	0.636
C	0.288	0.666	0.933	0.986	0.276	0.284	0.555	0.769
	$H = 0.8$				$H = 0.9$			
n	50	100	250	400	50	100	250	400
S	0.101	0.221	0.241	0.350	0.057	0.098	0.236	0.167
W	0.089	0.156	0.264	0.383	0.089	0.116	0.191	0.147
C	0.171	0.234	0.348	0.349	0.164	0.127	0.239	0.223

	Relative change position $\tau = 0.5$							
	$H = 0.6$				$H = 0.7$			
n	50	100	250	400	50	100	250	400
S	0.664	0.919	1.000	1.000	0.524	0.682	0.925	0.970
W	0.621	0.930	0.998	1.000	0.513	0.742	0.906	0.967
C	0.733	0.918	0.997	0.999	0.599	0.717	0.919	0.960
	$H = 0.8$				$H = 0.9$			
n	50	100	250	400	50	100	250	400
S	0.418	0.504	0.655	0.830	0.359	0.461	0.475	0.526
W	0.374	0.485	0.674	0.770	0.387	0.430	0.578	0.587
C	0.400	0.553	0.673	0.766	0.393	0.499	0.522	0.553

We then use the empirical quantiles of $\{S_{n,j}\}_{j=1}^{J}$ as critical values. The Cramér-von Mises statistic is invariant under monotone transformations of the data (as is the Kolmogorov-Smirnov statistic). Hence the critical values are valid if our observations are monotone transformations of Gaussian data. We note that this is a strong assumption and that an accurate approximation of the empirical process for general long-range dependent data is an issue of future research. The CUSUM statistic is not invariant under monotone transformations. Therefore, the Wilcoxon change-point test is considered additionally.

In the first part of the simulation study we treat realizations of a Gaussian process $X_1, \ldots, X_n$ given by *fractional Gaussian noise*). For the implementation we have used the function `fgnSim` from the R-package `fArma`. Eventually a change is added by $Y_i = X_i + \mu 1_{\{i > \lfloor n\tau \rfloor\}}$ and the three mentioned change-point tests are applied to $Y_1, \ldots, Y_n$.

Table 2: Empirical size, estimated Hurst coefficient.

	$H = 0.6$				$H = 0.7$			
n	50	100	250	400	50	100	250	400
$S_{\hat{H}}$	0.067	0.088	0.065	0.058	0.080	0.076	0.063	0.043
$S_{\hat{H}_{\hat{k}}}$	0.082	0.105	0.085	0.083	0.122	0.143	0.107	0.081
$W_{\hat{H}}$	0.067	0.058	0.054	0.057	0.081	0.074	0.061	0.035
$W_{\hat{H}_{\hat{k}}}$	0.070	0.100	0.102	0.078	0.122	0.118	0.100	0.081
$C_{\hat{H}}$	0.071	0.072	0.064	0.046	0.111	0.080	0.056	0.056
$C_{\hat{H}_{\hat{k}}}$	0.085	0.090	0.103	0.081	0.116	0.127	0.095	0.078
	$H = 0.8$				$H = 0.9$			
n	50	100	250	400	50	100	250	400
$S_{\hat{H}}$	0.094	0.104	0.062	0.063	0.075	0.070	0.092	0.080
$S_{\hat{H}_{\hat{k}}}$	0.136	0.141	0.099	0.077	0.087	0.127	0.118	0.112
$W_{\hat{H}}$	0.085	0.074	0.056	0.059	0.073	0.071	0.097	0.079
$W_{\hat{H}_{\hat{k}}}$	0.129	0.137	0.087	0.074	0.098	0.146	0.115	0.103
$C_{\hat{H}}$	0.165	0.101	0.051	0.046	0.309	0.257	0.112	0.075
$C_{\hat{H}_{\hat{k}}}$	0.218	0.127	0.094	0.072	0.137	0.130	0.083	0.063

If the Hurst-coefficient is assumed to be known, the empirical size of the tests naturally equals the nominal one, due to the construction of the critical values. The empirical power of Cramér-von Mises (denoted by S_n), Wilcoxon (denoted by W_n) and CUSUM test (denoted by C_n) is displayed in Table 1. If the change occurs in the middle of the observation period, the three tests are showing almost exactly the same performance, which matches the theoretical results. For early changes (after 20% of the observations) the CUSUM test is slightly more accurate than the other tests. Depending on sample size and strength of dependence, either the Cramér-von Mises or the Wilcoxon test might be second best.

3.2 Unknown Hurst coefficient

In applications the true Hurst coefficient H is unknown, and in the following we will consider two different estimators. The first is the local Whittle estimator (denoted by $\hat{H}$) with bandwidth parameter $m = \lfloor n^{2/3} \rfloor$, see Künsch (1987). However, if there is actually a change in the data, the local Whittle estimator is known to be biased. For the second estimator we therefore divide the observations into two subsamples

$$X_1, \ldots, X_{\hat{k}} \quad \text{and} \quad X_{\hat{k}+1}, \ldots, X_n$$

Table 3: Empirical Power, estimated Hurst coefficient, size of level shift $\mu = 1$, relative change positions $\tau = 0.2$ and $\tau = 0.5$.

	Relative change position $\tau = 0.2$							
	$H = 0.6$				$H = 0.7$			
n	50	100	250	400	50	100	250	400
$S_{\hat{H}}$	0.178	0.263	0.597	0.828	0.189	0.177	0.285	0.404
$S_{\hat{H}_{\hat{k}}}$	0.248	0.491	0.834	0.946	0.260	0.377	0.528	0.620
$W_{\hat{H}}$	0.190	0.298	0.625	0.811	0.146	0.193	0.278	0.386
$W_{\hat{H}_{\hat{k}}}$	0.291	0.532	0.850	0.943	0.237	0.364	0.488	0.649
$C_{\hat{H}}$	0.289	0.413	0.706	0.896	0.312	0.266	0.379	0.547
$C_{\hat{H}_{\hat{k}}}$	0.378	0.596	0.874	0.972	0.312	0.422	0.585	0.721
	$H = 0.8$				$H = 0.9$			
n	50	100	250	400	50	100	250	400
$S_{\hat{H}}$	0.115	0.125	0.156	0.185	0.096	0.105	0.142	0.163
$S_{\hat{H}_{\hat{k}}}$	0.208	0.265	0.287	0.296	0.160	0.213	0.208	0.234
$W_{\hat{H}}$	0.150	0.146	0.161	0.181	0.094	0.126	0.143	0.158
$W_{\hat{H}_{\hat{k}}}$	0.214	0.271	0.281	0.322	0.137	0.188	0.237	0.234
$C_{\hat{H}}$	0.405	0.321	0.270	0.305	0.539	0.442	0.367	0.328
$C_{\hat{H}_{\hat{k}}}$	0.313	0.321	0.383	0.429	0.418	0.350	0.311	0.313

	Relative change position $\tau = 0.5$							
	$H = 0.6$				$H = 0.7$			
n	50	100	250	400	50	100	250	400
$S_{\hat{H}}$	0.541	0.759	0.985	0.999	0.412	0.557	0.814	0.904
$S_{\hat{H}_{\hat{k}}}$	0.614	0.860	0.990	0.999	0.539	0.710	0.881	0.952
$W_{\hat{H}}$	0.594	0.811	0.984	0.999	0.441	0.564	0.809	0.902
$W_{\hat{H}_{\hat{k}}}$	0.609	0.877	0.991	1.000	0.550	0.717	0.878	0.950
$C_{\hat{H}}$	0.677	0.819	0.988	1.000	0.584	0.671	0.850	0.925
$C_{\hat{H}_{\hat{k}}}$	0.694	0.905	0.995	0.998	0.567	0.760	0.920	0.953
	$H = 0.8$				$H = 0.9$			
$S_{\hat{H}}$	0.373	0.403	0.535	0.640	0.337	0.428	0.472	0.549
$S_{\hat{H}_{\hat{k}}}$	0.454	0.563	0.649	0.711	0.412	0.480	0.582	0.604
$W_{\hat{H}}$	0.369	0.413	0.563	0.648	0.357	0.387	0.506	0.536
$W_{\hat{H}_{\hat{k}}}$	0.443	0.566	0.662	0.710	0.433	0.528	0.544	0.599
$C_{\hat{H}}$	0.576	0.584	0.655	0.706	0.659	0.622	0.631	0.637
$C_{\hat{H}_{\hat{k}}}$	0.535	0.599	0.716	0.760	0.604	0.574	0.637	0.638

Table 4: Empirical Power, estimated Hurst coefficient, relative change position $\tau = 0.5$, level shift of size $\mu = 1$ and change in variance from $\sigma^2 = 1$ to $\sigma_0^2 = 5/4$, nominal size $\alpha = 0.05$.

	$H = 0.6$				$H = 0.7$			
n	50	100	250	400	50	100	250	400
$S_{\hat{H}}$	0.701	0.931	1.000	1.000	0.606	0.772	0.963	0.996
$S_{\hat{H}_{\hat{k}}}$	0.878	0.986	1.000	1.000	0.817	0.973	0.999	1.000
$W_{\hat{H}}$	0.609	0.812	0.989	1.000	0.572	0.593	0.879	0.952
$W_{\hat{H}_{\hat{k}}}$	0.734	0.973	1.000	1.000	0.705	0.907	0.988	0.998
$C_{\hat{H}}$	0.660	0.899	0.999	1.000	0.529	0.724	0.938	0.983
$C_{\hat{H}_{\hat{k}}}$	0.588	0.916	1.000	1.000	0.507	0.806	0.983	0.998
	$H = 0.8$				$H = 0.9$			
$S_{\hat{H}}$	0.466	0.636	0.824	0.898	0.507	0.599	0.755	0.797
$S_{\hat{H}_{\hat{k}}}$	0.762	0.942	0.983	1.000	0.824	0.960	0.993	0.993
$W_{\hat{H}}$	0.597	0.568	0.669	0.727	0.718	0.645	0.582	0.634
$W_{\hat{H}_{\hat{k}}}$	0.616	0.850	0.944	0.981	0.562	0.823	0.926	0.953
$C_{\hat{H}}$	0.445	0.601	0.806	0.853	0.438	0.551	0.731	0.762
$C_{\hat{H}_{\hat{k}}}$	0.443	0.635	0.852	0.937	0.460	0.576	0.781	0.847

and estimate H on each set, using again the local Whittle estimator. Finally the new estimator is given by $\hat{H}_{\hat{k}} = \hat{k}/n\hat{H}_1 + (n-\hat{k})/n\hat{H}_2$. Here $\hat{k}$ is the natural change-point estimator, associated to each test. For example, in case of the Cramér-von Mises test we use

$$\hat{k} = \min\left\{ 1 \leq k \leq n - 1 \mid U_{k,n} = \max_{1 \leq k \leq n-1} U_{k,n} \right\},$$

where

$$U_{k,n} = \int_{x \in \mathbb{R}} \left(\sum_{i=1}^{k} 1_{\{X_i \leq x\}} - \frac{k}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}} \right)^2 \, d\hat{F}_n(x).$$

Consistency of this estimator was shown in Hariz et al. (2009). Horvath and Kokoszka (1997) verified consistency for the analogous CUSUM-based estimator.

Empirical size and empirical power of the tests under unknown H are displayed in tables 2 and 3. Let us first compare the impact of the different estimators $\hat{H}$ and $\hat{H}_{\hat{k}}$ on the finite sample performance of the Cramér-von Mises test. If we use the classical local Whittle estimator, the empirical size of the test is quite accurate and even matches the nominal size for $n = 400$ and $H \leq 0.8$. However, there is a loss in the empirical power. The power performance is much better, if the local Whittle estimator is modified. Actually there is no loss in power if compared to the case where H was assumed to be known. Then again, the probability of a false rejection is higher than $\alpha = 0.05$, so the test is quite liberal.

Table 5: Empirical Power, estimated Hurst coefficient, relative change position $\tau = 0.5$, $G_1(x) = x^2$, $G_2(x) = x^2 + x/2 + 1/2$, nominal size $\alpha = 0.05$, H is the Hurst coefficient of the underlying Gaussian.

	$H = 0.6$				$H = 0.7$			
n	50	100	250	400	50	100	250	400
$S_{\hat{H}}$	0.535	0.827	0.983	0.999	0.487	0.758	0.957	0.988
$S_{\hat{H}_{\hat{k}}}$	0.494	0.815	0.992	0.999	0.479	0.750	0.968	0.995
$W_{\hat{H}}$	0.480	0.743	0.990	0.998	0.430	0.632	0.924	0.986
$W_{\hat{H}_{\hat{k}}}$	0.420	0.735	0.986	1.000	0.407	0.657	0.933	0.987
$C_{\hat{H}}$	0.424	0.616	0.853	0.958	0.399	0.547	0.745	0.858
$C_{\hat{H}_{\hat{k}}}$	0.387	0.569	0.828	0.920	0.390	0.546	0.755	0.891
	$H = 0.8$				$H = 0.9$			
$S_{\hat{H}}$	0.461	0.670	0.825	0.885	0.424	0.522	0.607	0.614
$S_{\hat{H}_{\hat{k}}}$	0.474	0.680	0.860	0.934	0.507	0.589	0.698	0.704
$W_{\hat{H}}$	0.376	0.537	0.670	0.773	0.350	0.369	0.443	0.438
$W_{\hat{H}_{\hat{k}}}$	0.418	0.555	0.738	0.828	0.458	0.496	0.488	0.511
$C_{\hat{H}}$	0.374	0.464	0.564	0.596	0.384	0.352	0.343	0.352
$C_{\hat{H}_{\hat{k}}}$	0.440	0.538	0.655	0.772	0.397	0.491	0.562	0.579

Next we compare Cramér-von Mises, Wilcoxon and CUSUM test. The empirical size of the three tests is similar, no matter which estimator we choose and which situation we assume (sample size, Hurst coefficient), see Table 2.

In terms of empirical power the Cramér-von Mises and Wilcoxon test give similar results with the CUSUM test being slightly ahead for $\tau = 1/2$ and being clearly advantageous for early changes $\tau = 1/5$ (see Table 3).

We have to keep in mind that CUSUM and Wilcoxon test are designed to detect changes in the mean. On the contrary, the Cramér-von Mises test is a so called omnibus test and has power against arbitrary changes in the marginal distribution.

Therefore, we consider another situation, with the mean-shift being now accompanied by a small change in the variance. In detail,

$$
Y_i = \begin{cases} X_i & \text{for } i \leq \lfloor n\tau \rfloor, \\ \sigma X_i + \mu & \text{for } i > \lfloor n\tau \rfloor, \end{cases}
$$

for Gaussian $\{X_i\}_{i \geq 1}$. The theoretic result from Example 2.11 indicates that in this scenario the Cramér-von Mises test should be advantageous. In fact, for all combinations of sample size n and Hurst coefficient H the empirical power against this change is always higher than the power against a mean-shift under constant variance. Moreover, the Cramér-von Mises test has

Table 6: Empirical size and power for $farima(0, 0.2, 0)$-sequences, Hurst coefficient is estimated, nominal size $\alpha = 0.05$.

	No change				Mean-shift $\mu = 1$			
n	50	100	250	400	50	100	250	400
$S_{\hat{H}}$	0.036	0.058	0.069	0.056	0.276	0.490	0.832	0.920
$S_{\hat{H}_{\hat{k}}}$	0.167	0.148	0.119	0.129	0.520	0.737	0.939	0.986
$W_{\hat{H}}$	0.067	0.139	0.086	0.074	0.601	0.882	0.968	0.730
$W_{\hat{H}_{\hat{k}}}$	0.247	0.189	0.160	0.153	0.573	0.711	0.934	0.980
$C_{\hat{H}}$	0.104	0.061	0.053	0.048	0.281	0.479	0.836	0.945
$C_{\hat{H}_{\hat{k}}}$	0.212	0.202	0.158	0.114	0.499	0.682	0.937	0.977

clearly higher power then CUSUM and Wilcoxon test, which matches the theoretical findings of Example 2.11.

Moreover, we consider the change-point problem (based on non-monotone transformations)

$$
Y_i = \begin{cases} X_i^2 & \text{for } i \leq \lfloor n\tau \rfloor, \\ X_i^2 + aX_i + \mu & \text{for } i > \lfloor n\tau \rfloor, \end{cases}
$$

corresponding to a situation in which mean, variance, skewness and the Hermite rank change (see Example 2.8). Table 5 displays the empirical power of the three tests against this alternative and the picture is quite clear. The Cramér-von Mises test has the highest power for all combinations of H and n, while the Wilcoxon test is second best. Also note that the Hermite rank of the pre-change random variables is $m = 2$. Consequently, these observations are short-range dependent for $H < 0.75$.

3.3 $farima(0, d, 0)$-processes

For Gaussian long memory processes beyond fractional Gaussian noise, not only the Hurst coefficient determines the normalization. Instead it is given by

$$
d_n = n^H L^{1/2}(n)(H(2H - 1))^{1/2}),
$$

see (2.1). In this study we assume, as $n \to \infty$, $L(n) \to C$, which is quite common in the literature. For fractional Gaussian noise, $C = H(2H - 1)$ so the two factors just cancel out. In general the constant C is given through the limit

$$
\rho(k)k^{2-2H} \to C,
$$

Table 7: Empirical size and power for $farima(1, 0.2, 0)$-sequences with AR-coefficient $a_1 = 0.4$, Hurst coefficient is estimated, nominal size $\alpha = 0.05$.

	No change				Mean-shift $\mu = 1$			
n	50	100	250	400	50	100	250	400
$S_{\hat{H}}$	0.032	0.021	0.030	0.034	0.154	0.183	0.329	0.433
$S_{\hat{H}_{\hat{k}}}$	0.126	0.065	0.062	0.052	0.316	0.325	0.415	0.489
$W_{\hat{H}}$	0.038	0.009	0.003	0.007	0.157	0.158	0.192	0.266
$W_{\hat{H}_{\hat{k}}}$	0.183	0.048	0.015	0.017	0.363	0.313	0.301	0.387
$C_{\hat{H}}$	0.445	0.288	0.124	0.081	0.625	0.615	0.605	0.689
$C_{\hat{H}_{\hat{k}}}$	0.422	0.302	0.138	0.093	0.592	0.613	0.656	0.710

as $k \to \infty$. We suggest an estimator for C (which is quite heuristic) by:

$$\hat{C} = \frac{1}{K} \sum_{k=1}^{K} \hat{\rho}(k) k^{2-2\hat{H}}, \tag{3.1}$$

with $\hat{H}$ being one of the two estimators from above. Finally, we use the normalization

$$\hat{d}_n = n^{\hat{H}} \hat{C}^{1/2} (\hat{H}(2\hat{H} - 1))^{1/2}.$$

The estimator $\hat{C}$ in (3.1) is only defined under long memory, that is $H > 0.5$ (or in this situation $\hat{H} > 0.5$). Therefore, we modify both estimators by considering $\max(\hat{H}, 0.501)$ instead of $\hat{H}$. The effect of this modification on short memory processes will be seen in the next section.

However, for $farima(0, d, 0)$-sequences it seems to work quite well, see Table 6. Note that critical values are still deduced from fractional Gaussian noise. The finite sample performance (under the hypothesis as well as under a mean-shift) is very similar to the case where the data comes from fractional Gaussian noise. Meaning, the Cramér-von Mises test has good properties and the different tests yield very similar results, again matching the theoretic findings.

3.4 Short-range dependent effects

Finally, we have considered deviations from purely LRD sequences by simulating $farima(1, d, 0)$-time series and short memory AR(1)-processes.

First, we have applied the tests to $farima(0, d, 1)$-sequences, which are still long-range dependent. Table 7 indicates that the empirical power of the Cramér-von Mises test is less than in the case of $farima(0, d, 0)$-processes. However, the test works principally well, meaning that the power increases with the number of observations while the empirical size stays close to the nominal size. For CUSUM and Wilcoxon test this seems to be not the case.

Table 8: Empirical size and power for $AR(1)$-sequences with AR-coefficient $a_1 = 0.6$, Hurst coefficient is estimated, nominal size $\alpha = 0.05$.

	No change				Mean-shift $\mu = 1$			
n	50	100	250	400	50	100	250	400
$S_{\hat{H}}$	0.016	0.008	0.004	0.002	0.135	0.150	0.324	0.576
$S_{\hat{H}_{\hat{k}}}$	0.097	0.030	0.012	0.004	0.282	0.260	0.374	0.556
$W_{\hat{H}}$	0.036	0.004	0.000	0.000	0.149	0.106	0.106	0.213
$W_{\hat{H}_{\hat{k}}}$	0.141	0.018	0.000	0.000	0.336	0.249	0.235	0.343
$C_{\hat{H}}$	0.467	0.279	0.017	0.006	0.641	0.592	0.606	0.721
$C_{\hat{H}_{\hat{k}}}$	0.412	0.219	0.026	0.006	0.654	0.631	0.678	0.794

For the (purely short-range dependent) AR(1)-processes we make two observations: First, due to the assumption of LRD ($H > 0.5$) the normalization is too strong and the statistics converge to 0, at least under stationarity. If the structural change is big enough, the tests might still detect the change (see Table 8). However, there is a certain loss in power.

Secondly, Cramér-von Mises test and CUSUM test are showing a quite different finite sample performance. While under LRD (in concordance with the theory) their empirical size and power is always very similar, we now observe situations where the Cramér-von Mises test has empirical power 0.374 and the CUSUM test 0.678, see the results in Table 8. Again, this matches the theoretical fact that under short memory the tests show a different asymptotic behavior.

4 Proofs of the main results

4.1 Proof of Theorem 3 and Corollary 2.13

It is the goal to approximate the sequential empirical process by a linear combination of multiple partial sum processes. The indicator function $1_{\{G_n(X_j) \leq x\}}$ has the Hermite expansion

$$1_{\{G_n(X_j) \leq x\}} = \sum_{q=0}^{\infty} \frac{J_{q,n}(x)}{q!} H_q(X_j).$$

Remind that $J_{q,n}(x) = E[1_{\{G_n(X_j) \leq x\}} H_q(X_j)]$ and especially $J_{0,n}(x) = P(G_n(X_j) \leq x) = F_{(n)}(x)$. Now let $L_{m,n,j}(x)$ be the Hermite expansion up to m, in detail

$$L_{m,n,j}(x) = \sum_{q=0}^{m} \frac{J_{q,n}(x)}{q!} H_q(X_j). \tag{4.1}$$

Let $m(n)$ be the Hermite rank of $(1_{\{G_n(X_j)\leq x\}})_x$. Then we have by the conditions of Theorem 3 that $m^* \leq m(n) \leq m$ for some $m^* \leq m < 1/D$. Thus

$$L_{m,n,j}(x) = F_{(n)}(x) + \sum_{q=m^*}^{m} J_{q,n}(x)/q!H_q(X_j).$$

Moreover, define

$$S_n(l;x) = \frac{1}{d_{n,m}} \sum_{j=1}^{l} \left(1_{\{G_n(X_j)\leq x\}} - L_{m,n,j}(x) \right).$$

Finally, let $S_n(k;x,y) = S_n(k;y) - S_n(k;x)$, $L_{m,n,j}(x,y) = L_{m,n,j}(y) - L_{m,n,j}(x)$ and $J_{n,q}(x,y) = J_{n,q}(y) - J_{n,q}(x)$.

We will make use of the chaining technique of Dehling and Taqqu (1989a). To this end, define

$$\Lambda_n(x) := \int_{\{G_n(s)\leq x\}} \left(\sum_{q=0}^{m} \frac{|H_q(s)|}{q!} \right) \phi(s) \, ds$$

and observe that $J_{q,n}(x,y)/q!$ is bounded by $\Lambda_n(x,y) = \Lambda_n(y) - \Lambda_n(x)$, for all $n \in \mathbb{N}$ and all $q = 0,\dots,m$. Furthermore, Λ_n is monotone, $\Lambda_n(-\infty) = 0$ and

$$\Lambda_n(+\infty) = \int_{\mathbb{R}} \left(\sum_{q=0}^{m} \frac{|H_q(s)|}{q!} \right) \phi(s) \, ds = C < \infty, \quad \text{for all } n \in \mathbb{N}.$$

Define partitions, similarly to Dehling and Taqqu (1989a), but now depending on n, by

$$x_i(k) = x_i^{(n)}(k) = \inf\{x|\Lambda_n(x) \geq \Lambda_n(+\infty)i2^{-k}\} \quad i = 0,\cdots,2^k - 1$$

for $k = 0,\cdots,K$, with the integer K chosen below. Then we have

$$\Lambda_n(x_i(k)-) - \Lambda_n(x_{i-1}(k)) \leq \Lambda_n(+\infty)2^{-k}. \tag{4.2}$$

Note that the right hand side of (4.2) does not depend on n.

Based on these partitions we can define chaining points $i_k(x)$ by

$$x_{i_k(x)}(k) \leq x < x_{i_k(x)+1}(k),$$

for each x and each $k \in \{0,1,\dots,K\}$, see Dehling and Taqqu (1989a).

Lemma 4.1. *Define the chaining points as above. Suppose the following two conditions hold:*

27

(i) There are constants $\gamma > 0$ and $C > 0$, not depending on n, such that for all $k \leq n$

$$E|S_n(k; x, y)|^2 \leq C\left(\frac{k}{n}\right) n^{-\gamma} F_{(n)}(x, y).$$

(ii) For all $\epsilon > 0$ and all $n \in \mathbb{N}$ there is a real number $K = K(n, \epsilon)$, such that for all $\lambda > 0$

$$P\left(\sup_{x \in \mathbb{R}} \left| \frac{1}{d_{n,m}} \sum_{j=1}^{l} L_{m,n,j}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-) \right| > \epsilon\right) \leq C\left(\frac{l}{n}\right)^{2-m^*D} n^{\lambda - m^*D}.$$

Then there is a constant $\rho > 0$, such that for all $n \in \mathbb{N}$ and all $\epsilon > 0$ the following holds:

$$P\left(\sup_x |S_n(l; x)| > \epsilon\right) \leq C\left(\frac{l}{n}\right) n^{-\gamma} \epsilon^{-2} (K(n, \epsilon) + 3)^5 + C\left(\frac{l}{n}\right)^{2-m^*D} n^{\lambda - m^*D}.$$

Proof. Due to definition of the chaining points each point x is linked to $-\infty$ in detail

$$-\infty = x_{i_0(x)}(0) \leq x_{i_1(x)}(1) \leq \cdots \leq x_{i_K(x)}(K) \leq x < x_{i_K(x)+1}(K)$$

We have

$$S_n(l; x) = \sum_{k=1}^{K} S_n(l; x_{i_{k-1}(x)}(k-1), x_{i_k(x)}(k)) \ + \ S_n(l; x_{i_K(x)}(K), x). \tag{4.3}$$

The last summand of the right hand side of (4.3) can be treated as follows

$$
\begin{aligned}
\left| S_n(l; x_{i_K(x)}(K), x) \right| &= \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} \left(1_{\{x_{i_K(x)}(K) < G_n(X_j) \leq x\}} - L_{m,n,j}(x_{i_K(x)}(K), x) \right) \right| \\
&\leq \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} \left(1_{\{x_{i_K(x)}(K) < G_n(X_j) < x_{i_K(x)+1}(K)\}} - L_{m,n,j}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-) \right) \right| \\
&\quad + 2 \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} L_{m,n,j}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-) \right| \\
&= \left| S_n(l; x_{i_K(x)}(K), x_{i_K(x)+1}(K)-) \right| \\
&\quad + 2 \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} L_{m,n,j}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-) \right|.
\end{aligned}
$$

$$\tag{4.4}$$

By (4.3) and (4.4) we get, using $\sum_{k=1}^{\infty}(k+2)^{-2} < 1/2$,

$$P\left(\sup_x |S_n(l;x)| > \epsilon\right)$$

$$\leq P\left(\sup_x |S_n(l;x)| > \epsilon \sum_{k=1}^{K+1}(k+2)^{-2} + \epsilon/2\right)$$

$$\leq \sum_{k=1}^{K} P\left(\max_x |S_n(l;x_{i_{k-1}(x)}(k-1), x_{i_k(x)}(k))| > \epsilon/(k+2)^2\right) \qquad (4.5)$$

$$+ P\left(\max_x |S_n(l;x_{i_K(x)}(K), x_{i_K(x)+1}(K)-)| > \epsilon/(K+3)^2\right) \qquad (4.6)$$

$$+ P\left(2d_{n,m}^{-1}\left|\sum_{j\leq l} L_{m,n,j}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-)\right| > (\epsilon/2)\right). \qquad (4.7)$$

Further, by condition (i) of Lemma 4.1 and the Markov inequality we get

$$P\left(\max_x |S_n(l;x_{i_k(x)}(k), x_{i_{k+1}(x)}(k+1))| > \epsilon/(k+2)^2\right)$$

$$\leq \sum_{i=0}^{2^{k+1}-1} P\left(S_n(l;x_i(k+1), x_{i+1}(k+1)) > \epsilon/(k+2)^2\right)$$

$$\leq C \sum_{i=0}^{2^{k+1}-1} \left(\frac{l}{n}\right) n^{-\gamma}\frac{(k+2)^4}{\epsilon^2}F_{(n)}(x_i(k+1), x_{i+1}(k+1)) \qquad (4.8)$$

$$\leq C\left(\frac{l}{n}\right) n^{-\gamma}\frac{(k+2)^4}{\epsilon^2}.$$

The constant C in (4.8) is the constant of condition (i) in Lemma 4.1 and thus independent of n. In the next line this C gets multiplied with $\Lambda_n(+\infty)$, which is a constant by itself. Thus the C in the inequality above is a universal constant, not depending on n. The same is true for γ.

Using the same arguments we get moreover

$$P\left(\max_x |S_n(l;x_{i_K(x)}(K), x_{i_K(x)+1}(K)-)| > \epsilon/(K+3)^2\right) \leq C\left(\frac{l}{n}\right) n^{-\gamma}\frac{(K+3)^4}{\epsilon^2}.$$

Finally we have by condition (ii) of Lemma 4.1

$$P\left(2d_{n,m}^{-1}\left|\sum_{j\leq l} L_{m,n,j}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-)\right| > (\epsilon/2)\right) \leq C\left(\frac{l}{n}\right)^{2-m^*D} n^{\lambda-m^*D},$$

for all $\lambda > 0$. Combining the estimates for (4.5), (4.6) and (4.7) we arrive at

$$
P\left(\sup_x |S_n(l;x)| > \epsilon\right) \leq C\left(\frac{l}{n}\right) n^{-\gamma}\epsilon^{-2}\sum_{k=1}^{K+1}(k+2)^4 + C\left(\frac{l}{n}\right)^{2-m^*D} n^{\lambda-m^*D}
$$

$$
\leq C\left(\frac{l}{n}\right) n^{-\gamma}\epsilon^{-2}(K+3)^5 + C\left(\frac{l}{n}\right)^{2-m^*D} n^{\lambda-m^*D}.
$$

which finishes the proof. $\qquad\square$

Lemma 4.2. *There exist constants γ and C, not depending on n, such that for all $k \leq n$*

$$
E|S_n(k;x,y)|^2 \leq C\left(\frac{k}{n}\right) n^{-\gamma}F_{(n)}(x,y).
$$

The proof is very close to the proof of Lemma 3.1 in Dehling and Taqqu (1989a). However, for further results it is crucial that C and γ only depend indirectly on the function G_n, namely through the Hermite rank. Thus we give a detailed proof to highlight this fact.

Proof. First, obtain the Hermite expansion

$$
1_{\{x < G_n(X_i) \leq y\}} - F_{(n)}(x,y) = \sum_{q=m^*}^{\infty} \frac{J_{q,n}(x,y)}{q!} H_q(X_i).
$$

Secondly, we have by orthogonality of the $H_q(X_i)$ and $EH_q^2(X_i) = q!$

$$
\begin{aligned}
\sum_{q=m^*}^{\infty} \frac{J_{q,n}^2(x,y)}{q!} &= \sum_{q=m^*}^{\infty} E\left(\frac{J_{q,n}(x,y)}{q!} H_q(X_i)\right)^2 \\
&= E\left(\sum_{q=m^*}^{\infty} \frac{J_{q,n}(x,y)}{q!} H_q(X_i)\right)^2 \\
&= E\left(1_{\{x < G_n(X_i) \leq y\}} - F_{(n)}(x,y)\right)^2 \\
&= F_{(n)}(x,y)(1 - F_{(n)}(x,y)) \\
&\leq F_{(n)}(x,y).
\end{aligned}
$$

This yields

$$
\begin{aligned}
E\left(d_{n,m}S_n(k;x,y)\right)^2 &= \sum_{q=m+1}^{\infty} \frac{J_{q,n}^2(x)}{q!} \frac{1}{q!} \sum_{i,j \leq k} EH_q(X_i)H_q(X_j) \\
&\leq F_{(n)}(x,y) \sum_{i,j \leq k} |r(i-j)|^{m+1}.
\end{aligned}
$$

Note that the second factor of the product in the last line may depend indirectly on the function

G_n, because G_n determines m, however this is the only influence. For different combinations of m and D the term $\sum_{i,j \leq k} |r(i-j)|^{m+1}$ might have a different asymptotic order. However, in all cases we get (see page 1777 in Dehling and Taqqu (1989a))

$$\frac{1}{d_{n,m}} \sum_{i,j \leq k} |r(i-j)|^{m+1} \leq Cn^{mD-2} L^{-m}(n) k^{1 \vee (2-(m+1))/D} L_1(k)$$

$$\leq C \left(\frac{k}{n}\right)^{1 \vee (2-(m+1)D)} n^{mD-1 \vee (-D)} L_1(k) L^{-m}(n).$$

The result then follows because L and L_1 are slowly varying. □

Lemma 4.3. *Let $n \in \mathbb{N}$ and $\epsilon > 0$. Define the chaining points and $L_{m,n,j}(x)$ as in (4.1). Set*

$$K = K(n, \epsilon) = \left\lfloor \log_2 \left(\frac{(m - m^* + 2)\Lambda_n(+\infty)}{\epsilon} n d_{n,m}^{-1} \right) \right\rfloor + 1.$$

Then there is a constant $C > 0$, such that for all $\lambda > 0$

$$P\left(\sup_{x \in \mathbb{R}} \left| \frac{1}{d_{n,m}} \sum_{j=1}^{l} L_{m,n,j}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-) \right| > \epsilon \right) \leq C \left(\frac{l}{n}\right)^{2-m^*D} n^{\lambda - m^*D}.$$

Proof. By construction of the chaining points we have for $q = 0, \ldots, m$ and for all $x \in \mathbb{R}$

$$\sup_{x \in \mathbb{R}} |J_{q,n}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-)/q!| \leq \Lambda_n(+\infty) 2^{-K}.$$

Thus for all $x \in \mathbb{R}$

$$\frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} L_{m,n,j}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-) \right|$$

$$\leq \sum_{q=0}^{m} |J_{q,n}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-)/q!| \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} H_q(X_j) \right|$$

$$\leq \Lambda_n(+\infty) 2^{-K} \sum_{q=0}^{m} \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} H_q(X_j) \right|.$$

By definition of K

$$\frac{2^K \epsilon}{(m - m^* + 2)\Lambda_n(+\infty)} \geq \frac{n}{d_{n,m}}.$$

31

Therefore we get by Markov's inequality for $q = m^*, \ldots, m$

$$P\left(\Lambda_n(+\infty)2^{-K}\frac{1}{d_{n,m}}\left|\sum_{j=1}^{l}H_q(X_j)\right| > \epsilon/(m-m^*+2)\right) \leq P\left(\left|\sum_{j=1}^{l}H_q(X_j)\right| > n\right)$$

$$\leq C\frac{d_{l,q}^2}{n^2}$$

$$\leq C\frac{l^{2-qD}}{n^2}L^q(l)$$

$$\leq C\left(\frac{l}{n}\right)^{2-m^*D}n^{\lambda-m^*D}.$$

For $q = 0$ the term is deterministic, thus the probability is 0. $\qquad\square$

Proof of Theorem 3. The two conditions of Lemma 4.1 are satisfied (see Lemma 4.2 and Lemma 4.3) with

$$K = \left\lfloor \log_2\left(\frac{(m-m^*-2)\Lambda_n(+\infty)}{\epsilon}nd_{n,m}^{-1}\right)\right\rfloor + 1.$$

Note that $(K+3)^5 \leq C\epsilon^{-1}n^{\delta}$ for any $\delta > 0$, see Dehling and Taqqu (1989a), page 1781. By this fact and by virtue of Lemma 4.1

$$P\left(\sup_x|S_n(l;x)| > \epsilon\right) \leq C\left(\frac{l}{n}\right)n^{\delta-\gamma}\epsilon^{-3} + C\left(\frac{l}{n}\right)^{2-m^*D}n^{\lambda-m^*D}$$

$$\leq Cn^{-\rho}\left\{\left(\frac{l}{n}\right)\epsilon^{-3} + \left(\frac{l}{n}\right)^{2-m^*D}\right\},$$

with $\rho = \min(\gamma - \delta, m^*D - \lambda)$. Now choose $\delta < \gamma$, then $\rho > 0$ and we have thus proven a reduction principle in x. It remains to verify uniformity in l. For $n = 2^r$ one gets by the same arguments as in the proof of Theorem 3.1 in Dehling and Taqqu (1989a)

$$P\left(\max_{l\leq n}\sup_x|S_n(l;x)| > \epsilon\right) \leq Cn^{-\kappa}(1+\epsilon^{-3})$$

for any $0 < \epsilon \leq 1$ and universal constants C and ϵ. Next consider arbitrary n and define for r such that $2^{r-1} < n \leq 2^r$

$$S_n^*(l,x) = \frac{1}{d_{2^r,m}}\sum_{j=1}^{l}(1_{\{G_n(X_j)\leq x\}} - L_{m,n,j}(x)) \quad \text{for } l \leq 2^r,$$

where $\{G_n(X_j)\}_{n\in\mathbb{N}, j\leq 2^r}$ is a (slightly modified) array. One obtains

$$P\left(\max_{l\leq n}\sup_x|S_n^*(l;x)| > \epsilon\right) \leq C(2^r)^{-\kappa}(1+\epsilon^{-3}).$$

Hence

$$\begin{aligned}
P\left(\max_{l\leq n}\sup_x|S_n(l;x)| > \epsilon\right) &\leq P\left(\max_{l\leq n}\sup_x|S_n^*(l;x)| > \epsilon\frac{d_{n,m}}{d_{2^r,m}}\right)\\
&\leq\ \leq C(2^r)^{-\kappa}\left(1+\epsilon^{-3}\left(\frac{d_{2^r,m}}{d_{n,m}}\right)^3\right)\\
&\leq Cn^{-\kappa}(1+\epsilon^{-3}).
\end{aligned}$$

The last line holds since $d_{2^r,m}/d_{n,m}$ is uniformly bounded away from 0 and ∞. Thus, Theorem 3 is proven. $\qquad\square$

Proof of Corollary 2.13. Using the reduction principle, namely Theorem 3, it remains to show that

$$d_{n,m}^{-1}\sum_{q=m^*}^m \frac{J_{q,n}(x)}{q!}\sum_{i=1}^{\lfloor nt\rfloor} H_q(X_i) \tag{4.9}$$

converges to the desired limit processes. Define

$$Z_{n,q}(t) = \frac{1}{d_{n,q}}\sum_{i=1}^{\lfloor nt\rfloor} H_q(X_i),$$

and note that because of $1/m > D$ the sequences $\{H_q(X_i)\}_{i\geq 1}$ are long-range dependent for $q = m^*,\dots,m$. Then we have by Theorem 4 of Bai and Taqqu (2013)

$$(Z_{n,m^*},\dots,Z_{n,m}) \xrightarrow{\mathcal{D}} (Z_{m^*},\dots,Z_m), \tag{4.10}$$

where convergence takes place in $(D[0,1])^{m-m^*+1}$, equipped with the uniform metric. Moreover, $(Z_q(t))_{t\in[0,1]}$ are uncorrelated Hermite processes of order q. The functions h_q are elements of $D[-\infty,\infty]$ and therefore they are also bounded, see Remark 2.15. Hence we may apply the continuous mapping theorem and conclude that

$$\left\{\sum_{q=m^*}^m h_q(x)d_{n,q}^{-1}\sum_{i=1}^{\lfloor nt\rfloor} H_q(X_i)\right\}_{t,x}$$

converges in distribution to

$$\left\{ \sum_{q=m^*}^{m} h_q(x) Z_q(t) \right\}_{t,x},$$

where convergence takes place in $D([0,1] \times [-\infty, \infty]$, equipped with the supremum norm. The result then follows by the uniform convergence of $d_{n,m}/d_{n,q} J_{q,n}(x)$ towards $q! h_q(x)$, the reduction principle and Slutsky's theorem. $\qquad\square$

4.2   Proof of Theorem 1 and Theorem 2

We start by proving a reduction principle for the empirical process in presence of a change point. Consider the array $\{Y_{n,i}\}_{n \in \mathbb{N}, i \leq n}$, defined in section 2, and let $H_{n,i}(x) = P(Y_{n,i} \leq x)$. Define

$$S_n^{(\tau)}(t,x) = \frac{1}{d_{n,m}} \sum_{i=1}^{\lfloor nt \rfloor} \left(1_{\{Y_{n,i} \leq x\}} - H_{n,i}(x) - \sum_{q=m^*}^{m} \frac{J_{q,n,i}(x)}{q!} H_q(X_i) \right),$$

where $J_{q,n,i}(x) = E[1_{\{Y_{n,i} \leq x\}} H_q(X_i)]$. Note that $J_{q,n,i}(x) = 0$ if $i \leq \lfloor n\tau \rfloor$ and $q < m$.

Lemma 4.4. *Let the conditions of Theorem 1 hold. Then there are constants $C \geq 0$ and $\kappa > 0$ such that for all $\epsilon > 0$*

$$P\left(\sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |S_n^{(\tau)}(t,x)| > \epsilon \right) \leq Cn^{-\kappa}(1 + \epsilon^{-3}).$$

Proof. Define

$$S_{n,1}(t,x) = \frac{1}{d_{n,m}} \sum_{j=1}^{\lfloor nt \rfloor} \left(1_{\{G(X_j) \leq x\}} - F(x) - \frac{J_m(x)}{m!} H_m(X_j) \right)$$

and $\quad S_{n,2}(t,x) = \frac{1}{d_{n,m}} \sum_{j=1}^{\lfloor nt \rfloor} \left(1_{\{G_n(X_j) \leq x\}} - F_{(n)}(x) - \sum_{q=m^*}^{m} \frac{J_{q,n}(x)}{q!} H_q(X_j) \right).$

By Theorem 3 we have

$$P\left(\sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |S_{n,i}(t,x)| > \epsilon \right) \leq Cn^{-\kappa}(1 + \epsilon^{-3}) \quad i = 1, 2. \qquad (4.11)$$

34

Next obtain

$$S_n^{(\tau)}(t, x) = \begin{cases} S_{n,1}(t, x), & \text{if } t \leq \tau, \\ S_{n,2}(t, x) + S_{n,1}(\tau, x) - S_{n,2}(\tau, x), & \text{if } t > \tau. \end{cases}$$

Therefore, we get, using (4.11) several times,

$$P\left(\sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |S_n^{(\tau)}(t, x)| > \epsilon\right) \leq 2P\left(\sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |S_{n,1}(t, x)| > \epsilon/4\right)$$

$$+ 2P\left(\sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |S_{n,2}(t, x)| > \epsilon/4\right)$$

$$\leq 4Cn^{-\kappa}(1 + \epsilon^{-3}),$$

for all $n \in \mathbb{N}$ and all $\epsilon > 0$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Lemma 4.5. *Let Assumption A hold. Then for all $q \leq m$*

$$\sup_{x \in \mathbb{R}} d_{n,m^*}/d_{n,m} |J_{q,n}(x) - J_q(x)| \to 0, \qquad\qquad (4.12)$$

as $n \to \infty$.

Proof. Using Hölder's inequality, one has for any $p \in \mathbb{N}$

$$|J_{q,n}(x) - J_q(x)| = |E\left((1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}})H_q(X_i)\right)|$$

$$\leq \left(E|1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}}|^{(p+1)/p}\right)^{p/(p+1)} \|H_q(X_i)\|_{L^{p+1}}$$

$$\leq C(E|1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}}|)^{p/(p+1)}$$

Now obtain

$$E|1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}}|$$

$$= P\left(\{G_n(X_1) \leq x, G(X_1) > x\} \cup \{G_n(X_1) > x, G(X_1) \leq x\}\right)$$

$$= 1 - P\left(\{G_n(X_1) \leq x, G(X_1) \leq x\}\right) - P\left(\{G_n(X_1) > x, G(X_1) > x\}\right)$$

$$= P(\min\{G_n(X_1), G(X_1)\} \leq x) - P(\max\{G_n(X_1), G(X_1)\} \leq x)$$

$$= o(n^{(m^*-m)D(1+\delta)/2}),$$

for some $\delta > 0$. The last line holds uniformly due to Assumptions A2. Finally,

$$
\begin{aligned}
& d_{n,m^*}/d_{n,m}|J_{q,n}(x) - J_q(x)| \\
& \leq Cn^{(m-m^*)D/2}L^{(m^*-m)/2}(n)(E|1_{\{G_n(X_i)\leq x\}} - 1_{\{G(X_i)\leq x\}}|)^{p/(p+1)} \\
& \leq C\left(n^{(m-m^*)D(p+1)/p/2}E|1_{\{G_n(X_i)\leq x\}} - 1_{\{G(X_i)\leq x\}}|\right)^{p/(p+1)} \\
& = C\left(n^{(m-m^*)D(1+1/p)/2}E|1_{\{G_n(X_i)\leq x\}} - 1_{\{G(X_i)\leq x\}}|\right)^{p/(p+1)}.
\end{aligned}
$$

Choosing $p > 1/\delta$, this implies (4.12). $\qquad\square$

Proof of Theorem 1. By definition of the functions $J_{q,n,i}$ we get

$$
\begin{aligned}
& \frac{1}{d_{n,m}}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{q=m^*}^{m}\frac{J_{q,n,i}(x)}{q!}H_q(X_i) \\
& = \frac{1}{d_{n,m}}\frac{J_m(x)}{m!}\sum_{i=1}^{\lfloor nt\rfloor}H_m(X_i) \\
& \quad + 1_{\{t>\tau\}}\sum_{q=m^*}^{m-1}\frac{d_{n,q}}{d_{n,m}}\frac{J_{q,n}(x)}{q!}\frac{1}{d_{n,q}}\sum_{i=\lfloor n\tau\rfloor+1}^{\lfloor nt\rfloor}H_q(X_i) \\
& \quad + 1_{\{t>\tau\}}\frac{J_{m,n}(x) - J_m(x)}{m!}\frac{1}{d_{n,m}}\sum_{i=\lfloor n\tau\rfloor+1}^{\lfloor nt\rfloor}H_m(X_i).
\end{aligned}
$$

The second and the third summands are negligible due to the uniform convergence of the functions $J_{q,n}$ (see Lemma 4.5)). The first summand converges in distribution towards

$$
\frac{J_m(x)}{m!}Z_m(t),
$$

see Dehling and Taqqu (1989a). Together with Lemma 4.4 this finishes the proof. $\qquad\square$

Proof of Theorem 2. We give the proof for a sequence of local alternatives. The asymptotic behavior under the hypothesis then is an immediate consequence. Obtain the following de-

composition of the empirical bridge-process

$$\frac{1}{d_{n,m}} \left(\sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right) \tag{4.13}$$

$$= \frac{1}{d_{n,m}} \left(\sum_{i=1}^{\lfloor nt \rfloor} \left(1_{\{Y_{n,i} \leq x\}} - H_{n,i}(x) \right) - t \sum_{i=1}^{n} \left(1_{\{Y_{n,i} \leq x\}} - H_{n,i}(x) \right) \right)$$

$$+ \left(t - \frac{\lfloor nt \rfloor}{n} \right) \frac{1}{d_{n,m}} \sum_{i=1}^{n} \left(1_{\{Y_{n,i} \leq x\}} - H_{n,i}(x) \right)$$

$$+ \frac{n}{d_{n,m}} \psi_{n,\tau}(t) \left(F(x) - F_{(n)}(x) \right),$$

where

$$\psi_{n,\tau}(t) = \begin{cases} \frac{\lfloor nt \rfloor}{n} \left(1 - \frac{\lfloor n\tau \rfloor}{n} \right), & \text{if } t \leq \tau, \\ \frac{\lfloor n\tau \rfloor}{n} \left(1 - \frac{\lfloor nt \rfloor}{n} \right), & \text{if } t > \tau. \end{cases}$$

By uniform convergence of $n/d_{n,m}(F(x) - F_{(n)}(x))$ and $\psi_{n,\tau}(t)$ towards $g(x)$ and $\psi_\tau(t)$, respectively, Theorem 1 and the continuous mapping theorem, one gets that (4.13) converges weakly towards

$$J_m(x)/(m!) \left(Z_m(t) - tZ_m(t) \right) + \psi_\tau(t)g(x).$$

The convergence of the Kolmogorov-Smirnov type statistic then follows from continuity of the application of the supremum norm. The Cramér-von Mises statistic S_n can be written $S_n = \sup_{t \in [0,1]} M_n(t)$, where

$$M_n(t) = d_{n,m}^{-2} \int_{\mathbb{R}} \left(\sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 d\hat{F}_n(x)$$

$$= d_{n,m}^{-2} \int_{\mathbb{R}} \left(\sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 dF(x) \tag{4.14}$$

$$+ d_{n,m}^{-2} \int_{\mathbb{R}} \left(\sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 d(\hat{F}_n(x) - F(x)). \tag{4.15}$$

Due to the convergence of (4.13) and the continuous mapping theorem, (4.14) converges to the desired limit process. Thus, it remains to show that (4.15) is negligible. Therefore, obtain

$$d_{n,m}^{-2} \int_{\mathbb{R}} \left(\sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 d(\hat{F}_n(x) - F(x))$$

$$= \int_{\mathbb{R}} \left(J_m(x)/(m!)\tilde{Z}_m(t) - \psi_\tau(t)g(x) \right)^2 d(\hat{F}_n(x) - F(x)) \tag{4.16}$$

$$+ \int_{\mathbb{R}} \left\{ d_{n,m}^{-2} \left(\sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 \right.$$

$$\left. - \left(J_m(x)/(m!)\tilde{Z}_m(t) - \psi_\tau(t)g(x) \right)^2 \right\} d(\hat{F}_n(x) - F(x)). \tag{4.17}$$

Using the Skorohod-Dudley-Wichura representation theorem (whose conditions are satisfied because $J_m(x)/(m!)\tilde{Z}_m(t) - \psi_\tau(t)g(x)$ lays almost surely in $C([0,1] \times [-\infty, \infty])$), one can assume without loss of generality that

$$d_{n,m}^{-2} \left(\sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 - \left(J_m(x)/(m!)\tilde{Z}_m(t) - \psi_\tau(t)g(x) \right)^2$$

converges almost surely to 0, uniformly in x and t. Thus, (4.17) converges to 0, uniformly in t. Next consider (4.16)

$$\int_{\mathbb{R}} \left(J_m(x)/(m!)\tilde{Z}_m(t) - \psi_\tau(t)g(x) \right)^2 d(\hat{F}_n(x) - F(x))$$

$$= (\tilde{Z}_m(t))^2/(m!)^2 \int_{\mathbb{R}} J_m^2(x) \, d(\hat{F}_n(x) - F(x))$$

$$- 2\tilde{Z}_m(t)\psi_\tau(t)/(m!) \int_{\mathbb{R}} J_m(x)g(x) \, d(\hat{F}_n(x) - F(x))$$

$$+ \psi_\tau^2(t) \int_{\mathbb{R}} g^2(x) \, d(\hat{F}_n(x) - F(x))$$

$$= I_n - II_n + III_n.$$

As a consequence of Theorem 1 and $F_{(n)}(x) \to F(x)$ one gets a weak Glivenko-Cantelli type

convergence, in detail

$$\sup_{x \in \mathbb{R}}|\hat{F}_n(x) - F(x)| \leq \sup_{x \in \mathbb{R}}\left|\hat{F}_n(x) - \sum_{i=1}^{n} \frac{H_{n,i}(x)}{n}\right| + \frac{n - \lfloor n\tau \rfloor}{n}\sup_{x \in \mathbb{R}}|F_{(n)}(x) - F(x)| \xrightarrow{P} 0.$$

Moreover, obtain that $J_m(x)$ is of bounded variation (this was also noted in Dehling and Taqqu (1989a)). To see this, let $[a, b]$ be an arbitrary interval and $\{x_i\}_{i=0}^{n}$ a partition of this interval. Then

$$
\begin{aligned}
\sum_{i=0}^{n-1}|J(x_{i+1}) - J(x_i)| &= \sum_{i=0}^{n-1}|E[1_{\{x_i < G(X_1) \leq x_{i+1}\}}H_m(X_1)]| \\
&\leq \sum_{i=0}^{n-1}E[1_{\{x_i < G(X_1) \leq x_{i+1}\}}|H_m(X_1)|] \\
&= E\left[\sum_{i=0}^{n-1}1_{\{x_i < G(X_1) \leq x_{i+1}\}}|H_m(X_1)|\right] \\
&= E\left[1_{\{G(X_1) \in [a,b]\}}|H_m(X_1)|\right] \\
&\leq E|H_m(X_1)|.
\end{aligned}
$$

By the boundedness of J_m, J_m^2 is also of bounded variation and thus integration by parts, together with the weak Glivenko-Cantelli-type reuslt, yields

$$I_n = -(\tilde{Z}_m(t))^2/(m!)^2 \int_{\mathbb{R}} (\hat{F}_n(x) - F(x)) \, dJ_m^2(x) \xrightarrow{P} 0.$$

By definition, the function $g(x)$ is bounded and of bounded variation. Hence the same is true for $g^2(x)$ and by the same arguments as above one gets $III_n = o_P(1)$. Finally, $II_n = o_P(1)$, which can be seen using Hölders's inequality. This finishes the proof. $\qquad\square$

Remark 4.6. Note that our proof of the weak convergence of the Cramér-von Mises statistic would not work for short-range dependent time series. The reason is the completely different limit behavior of the sequential empirical process. Instead of the semi-degenerate process $J_m(x)Z_m(t)$ one gets a Gaussian process $K(t, x)$. While J_m is of bounded variation, this is not the case for sample paths of K. Hence $\int_{\mathbb{R}} K(t, x) \, d(F_n(x) - F(x))$ cannot be treated simultaneously to $\int_{\mathbb{R}} J_m(x) \, d(F_n(x) - F(x))$.

Acknowledgements

References

C. Baek and V. Pipiras. Statistical test for a single change in the mean against long-range dependence. *J. Time Series Anal.*, 33:131–151, 2011.

S. Bai and M. Taqqu. Multivariate limit theorems in the context of long-range dependence. *J. Time Series Anal.*, 34:717–743, 2013.

J. Beran and N. Terrin. Testing for a change in the long-memory parameter. *Biometrika*, 83:627–638, 1996.

J. Beran, Y. Feng, S. Ghosh, and R. Kulik. *Long-Memory Processes*. Springer, 2013.

István Berkes, L. Horvath, P. Kokoszka, and Q. M. Shao. On discriminating between long-range dependence and changes in mean. *Ann. Statist.*, 34:1140–1165, 2006.

A. Betken. Testing for change-points in long-range dependent time series by means of a self-normalized wilcoxon test. *J. Time Series Anal.*, 37:785–809, 2016.

M. Csörgő and L. Horvath. *Limit Theorems in in Change-Point Analysis*. Wiley, 1997.

S. Csörgő and J. Mielniczuk. The empirical process of a short-range dependent stationary sequence under gaussian subordination. *Probab. Theory Related Fields*, 104:15–25, 1996.

H. Dehling and M. Taqqu. The empirical process of some long-range dependent data with application to U-statistics. *Ann. Statist.*, 17:1767–1783, 1989a.

H. Dehling and M. Taqqu. The functional law of the iterated logarithm for the empirical process of some long-range dependent sequences. *Statist. Probab. Lett.*, 7:81–85, 1989b.

H. Dehling, A. Rooch, and M. Taqqu. Nonparametric change-point tests for long-range dependent data. *Scand. J. Stat.*, 2012.

H. Dehling, A. Rooch, and M. Taqqu. Power of change-point tests for long-range dependent data. arXiv:1303.4917, 2013.

L. Giraitis, H. L. Koul, and D. Surgailis. Asymptotic normality of regression estimators with long memory errors. *Statist. Probab. Lett.*, 29:317–335, 1996a.

L. Giraitis, R. Leipus, and D. Surgailis. The change-point problem for dependent observations. *J. Statist. Plann. Inference*, 53:297–310, 1996b.

S.B. Hariz, W. Nie, J.Wylie, and Q. Zhang. Change-point detection for long-range dependent sequences in a general setting. *Nonlinear Anal.*, 71:2398–2405, December 2009.

H.C. Ho and T. Hsing. On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.*, 24:992–1024, 1996.

L. Horvath and P. Kokoszka. The effect of long-range dependence on change-point estimators. *J. Statist. Plann. Inference*, 64:57–81, 1997.

L. Horvath and Q. M. Shao. Limit theorems for quadratic forms with applications to whittle's estimate. *J. Appl. Probab.*, 9:146–187, 1999.

A. Inoue. Testing for distributional change in time series. *Econometric Theory*, 2001.

P. Kokoszka and R. Leipus. Detection and estimation of changes in regime. In P. Doukhan, G. Oppenheim, and M. Taqqu, editors, *Long-Range Dependence: Theory and Applications*. Birkhäuser, Boston, 2001.

H. Künsch. Statistical aspects of self-similar processes. In *Proc. First World Congress of the Bernoulli Society*, volume 1, pages 67–74, VNU Science Press,Utrecht, 1987. Yu. Prokhorov and V. V. Sazonov.

Ling. Testing for change points in time series models and limiting theorems for ned sequences. *Ann. Statist.*, 35:1213–1237, 2007.

G. E. Noether. Asymptotic properties of the Wals-Wolfowitz test of randomness. *Ann. Math. Statist.*, 21:231–246, 1950.

G.E. Noether. On a theorem of Pitman. *Ann. Math. Statist.*, 26:64–68, 1955.

E.J.G. Pitman. Unpuplished lecture notes. Columbia Univ., 1948.

D. Pollard. *Convergence of Stochastic Processes*. Springer, New York, 1984.

X. Shao. A simple test of changes in mean in the possible presence of longrange dependence. *J. Time Series Anal.*, 32:589–606, 2011.

Barbara Szyszkowicz. Weak convergence of weighted empirical type processes under contiguous and changepoint alternatives. *Stochastic Process. Appl.*, 50:281–313, April 1994.

M. Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete*, 31:287–302, 1975.

M. Taqqu. Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrsch. Verw. Gebiete*, 50:53–83, 1979.

J. Tewes. Block bootstrap for the empirical process of long-range dependent data. Preprint, arXiv:1601.01122, 2016.

W. B. Wu. Empirical processes of long-memory sequences. *Bernoulli*, 9:809–831, 2003.